Evolution: Naturalism's Kryptonite?

By Charl Louis Botha

Tesis ingelewer vir die graad Meester in die Wysbegeerte in die Fakulteit Lettere en Sosiale Wetenskappe aan die Universiteit van Stellenbosch.

Thesis presented for the degree of Master of Philosophy in the Faculty of Arts and Social Sciences at Stellenbosch University.

Supervisor: Prof. J.P. Simit

March 2021

Verklaring

Deur hierdie proefskrif elektronies in te lewer, verklaar ek dat die geheel van die werk hierin vervat, my eie, oorspronklike werk is, dat ek die alleenouteur daarvan is (behalwe in die mate uitdruklik anders aangedui), dat reproduksie en publikasie daarvan deur die Universiteit van Stellenbosch nie derdepartyregte sal skend nie en dat ek dit nie vantevore, in die geheel of gedeeltelik, ter verkryging van enige kwalifikasie aangebied het nie.

Declaration

By submitting this dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

March 2021

ABSTRACT

In this thesis, I consider two questions. The first is whether human cognitive faculties would likely be reliable on naturalism and evolution. The second is which of evolutionary naturalism or theistic evolution offers the best explanation of the empirical evidence we have with respect to human cognitive reliability.

With respect to the first question, Alvin Plantinga claims that human cognition would likely be unreliable given naturalism and evolution. For he claims that evolution's primary care is for creature's survival and not the truth-value of their beliefs. And that there is no naturalistic scenario on which the link between belief and behavior would be such that evolution can select for belief content, and where it can, no reason to think that it would select for mostly true content. Naturalist's disagree. In response, they argue that evolution would be able to select for belief content. And not only that, it would likely select for true content as well. For, according to them, it's plausible to think that true beliefs would on average be better guides to behaviour than false ones. But even if evolution can't select for belief content, a number of naturalists argue that there would be conceptual constraints on which beliefs could coherently be ascribed to a person's behaviour given her desires. And that this means that the most coherent belief ascriptions would involve mostly true beliefs.

With respect to this first question, I conclude that the naturalist's case is the more persuasive of the two. Human cognition would likely be reliable on naturalism and evolution.

With respect to the second question, I argue that evolutionary naturalism is the better explanation of the relevant empirical evidence we have concerning human cognitive reliability. For on Christian theism, it's of absolutely vital importance that humans know that the Christian God exists. That the evidence indicates that the majority of humanity don't know (or believe) that Christianity is true, and haven't for most of their history, is therefore very surprising. It is ultimately this that leads me to the conclusion that evolutionary naturalism offers the more plausible account of the evidence we have with respect to human cognitive reliability.

ABSTRAK

In hierdie tesis behandel ek twee vrae. Die eerste is of dit waarskynlik sal wees dat die mens se kognitiewe fakulteite betroubaar sal wees indien naturalisme en evolusie waar sou wees. Die tweede vraag is watter een van naturalistiese evolusie of tëistiese evolusie die beste verduideliking bied van die empiriese bewystukke wat ons het rakende die betroubaarheid van die mense se kognisie.

Rakende die eerste vraag beweer Alvin Plantinga dat die mens se kognitiewe fakulteite waarskynlik onbetroubaar sal wees indien naturalisme en evolusie waar sou wees. Hy maak die stelling dat evolusie primêr geïnteresseerd is in ´n wese se oorlewing en nie die waarheidswaarde van sy oortuigings nie. En dat daar geen scenario is waar die verwantskap tussen ´n wese se oortuings en dié se gedrag van so ´n aard is dat evolusie daarvoor kan selekteer nie, en waar dit kan, daar geen rede is om te glo dat dit meestal vir ware oortuigings sal selekteer nie. Naturaliste stem nie saam nie. In reaksie argumenteer hulle dat evolusie wél vir die oortuigings van ´n wese kan selekteer. En nie net dit nie, maar dat dit ook meestal vir ware oortuigings sal selekteer. Want volgens hulle is dit geloofwaardig om te dink dat ware oortuingings meerendeels meer geneig sal wees om tot suksesvolle gedrag te lei as wat valshede sou. Maar selfs as evolusie nie in staat is om vir oortuigings te selekteer nie, redeneer ´n aantal naturaliste dat konseptuele beperkinge sal geld rakende watter tipe oortuigings samehangend toegeskryf kan word tot ´n persoon se gedrag, gegewe haar begeertes. En dit sal beteken dat die mees samehangende geloofstoeskrywings meestal ware oortuingings sal insluit.

Met betrekking tot die eerste vraag vind ek die naturalis se argument die meer oortuigende van die twee is. Die mens se kognisie sal waarskynlik betroubaar wees indien naturalisme en evolusie waar sou wees.

Met betrekking tot die tweede vraag argumenteer ek dat naturalistiese evolusie ´n beter verduideling bied van die empiriese bewystukke wat ons het rakende die betroubaarheid van die mense se kognisie. Want volgens Christelike tëisme is dit van kardinale belang dat die mens weet dat die God van die Christendom bestaan. Dat die bewystukke wat ons tot ons beskikking het aandui dat die meerderheid van die mensdom nie kennis dra dat Christendom waar is nie (of nie glo dat dit so is nie), en dat dit gegeld het vir meeste van hul bestaan, is hoogs onverwags. Dit is uiteindelik wat my lei tot die konklusie dat naturalistiese evolusie ´n meer oortuigende verduideling bied van die empiriese bewystukke wat ons het rakende die betroubaarheid van die mense se kognisie.

ACKNOWLEDGMENTS

Firstly, I would like to thank my family, the Botha clan, for their moral and financial support. Without them, I certainly wouldn't have completed this thesis.

And secondly, I would like to thank my supervisor, Prof. J.P. Smit, for his patient guidance.
Stellenbosch, October 16, 2020

CONTENTS

INTRODUCTION

The theory of evolution is one of the key ideas on which a plausible and rational naturalism turns, where naturalism is roughly the idea that 'there is no such being as the God of traditional theism (or anyone like him)' (Plantinga, 1993: 220).[1] As renowned atheist Richard Dawkins has remarked: 'Darwin (has) made it possible to be an intellectually fulfilled atheist' (Dawkins, 1986: 6). On the contrary, argues philosopher Alvin Plantinga, far from enjoying its intellectual comforts, Dawkins and his cohorts should find Darwin's theory particularly disturbing. For he claims that the naturalist who is apprised of – and reflectively considers – his evolutionary argument against naturalism (the EAAN), but maintains that human cognition evolved, will be irrational for so doing.[2] In other words, according to Plantinga, the naturalist's position would ultimately be self-defeating. To remain epistemically respectable, she will have to forego either her naturalism or her evolution. But if Plantinga is right in thinking that evolution is arguably the 'only game in town' available to the naturalist in accounting for the origins of her cognitive faculties, then it appears as if it is her naturalism that would have to go. The Christian theist, on the other hand, would not find herself in a similar epistemic quandary. For, according to Plantinga, she has good reason to believe that God would likely have guided the evolutionary process such that the result was an epistemically happy one.

Plantinga's first step in his anti-naturalist project – and the subject matter of part 1 of this thesis – is the claim that the probability of the reliability of human cognition would be low on naturalism and evolution. In other words, if naturalism and evolution were both true, human cognition would likely be unreliable. The central idea on which Plantinga builds his case is the claim that evolution's primary care is for creatures reproductive success, not the truth-value of their beliefs. What ultimately matters is not what creatures *believe*, but how they *behave*. And he argues that there's no plausible naturalistic explanation on which belief – *as* belief – *can* be selected for, and where it *could*, no good reason to think that evolution would mostly select for *true* content. As I will show, naturalists have attempted to undermine or rebut premise one of Plantinga's argument in a number of different ways, some which I find persuasive and others less so.

As far as the evolutionary argument against naturalism is concerned, the primary reason for limiting myself to a discussion of premise 1 only is my belief that the arguments I will consider in response to it carry some force, (perhaps) enough to defeat it. If so, there wouldn't be any pressing need to evaluate the merits of Plantinga's greater anti-naturalist project. For if premise 1 is false it wouldn't matter whether every other premise of the

[1] Note that this is Plantinga's "definition" of naturalism as he employs it in the evolutionary argument against naturalism. As you can see, he views naturalism as a metaphysical doctrine – i.e. about what ultimately exists (or doesn't). The question as to the nature of naturalism – what it is and whether it is best seen as a metaphysical doctrine at all – is considered in detail in Chapter 2. This question is relevant to this work because if naturalism cannot most plausibly be conceived as a metaphysical view, then no argument which requires it to be seen in such a light – as the evolutionary argument against naturalism does – will work.

[2] It is important to be aware that the evolutionary argument against naturalism is not for *falsity* of naturalism, but its *rationality*. Indeed, if the evolutionary argument against naturalism is successful, it means that *even if naturalism is true*, it wouldn't be rational for anyone to believe it.

argument is true, the soundness of the argument would be compromised. The second reason is one of space constraint. As Beilby (2002) aptly remarks, Plantinga's argument is multi-layered and complex, and I suspect my efforts would have done it an injustice had I considered it in its entirety. And if I did manage to give it its due, I'm sure that it would have required a work much longer than the one you have before you.

Having said this, as I laboured on premise 1, I realised that there's an interesting – and closely related – question lurking in its shadow. And it is this: 'Which of evolutionary naturalism or theistic evolution best explains the evidence with respect to the reliability of human cognition in *this* world?' In other words, I wanted to consider not what *would* likely be true with respect to the reliability of human cognition – if either of evolutionary naturalism or theistic evolution were true – (as Plantinga does in premise 1), but which of the two best explains what is *in fact* the case. This question is essentially the inverse of premise. It's not: 'What would the probability of the reliability of human cognition be if naturalism and evolution were true?', but rather: 'How likely is *naturalism and evolution* given the evidence we have with respect to human cognitive reliability?' And how would the latter's likelihood compare to *theism* and evolution (given the same evidence)?

The importance of this "inverse" question to this project is *not* that it counters premise 1, but that if it turns out that evolutionary naturalism is the more likely of the two, there would be good reason to believe that this world is more likely fundamentally naturalistic than theistic (when all else is taken to be equal). And *this* would be important because if one can independently show that premise 1 is likely false, then evolution, far from being naturalism's kryptonite, is in fact one of its (relative) strengths. The question as to which of evolutionary naturalism or theistic evolution best explains the observed profile of human cognitive reliability is the subject matter of part 2 of this work.

In short, the aim of this thesis is thus two-fold; the first is to show that premise 1 of the evolutionary argument against naturalism is false. And the second is to consider which of evolutionary naturalism or theistic evolution best explains the *de facto* evidence we have with respect to the reliability of human cognition. If the first aim can be achieved, then evolution wouldn't be naturalism's kryptonite, for there would be no reason – given only Plantinga's evolutionary argument naturalism argument – to think that human cognition would likely be unreliable on naturalism and evolution. Indeed, if evolutionary naturalism proves to be the better explanation of the evidence we have with respect to (the profile) of human cognitive reliability, then evolution, far from being naturalism's kryptonite, would in fact provide some of the most persuasive evidence in its favour (at least vis-à-vis its theistic counterpart). In what follows, I will outline how I intend to achieve these twin aims.

But before I do so, I think it's worthwhile to highlight how premise 1 fits into the bigger evolutionary-argument-picture, if only to highlight Plantinga's ingenuity (and verve if I might add). For, at first sight, challenging naturalism by way of *evolution* wouldn't appear to be a very promising strategy. That his argument has drawn – and continues to draw – much critical attention is testimony to its creativity and probative force.[3]

[3] The most notable example probably being Beilby's (2002) volume specifically dedicated to such attempted rebuttals. For a sampling of more recent examples, see Childers (2011), Law (2011), Law (2012), and Boudry & Vlerick (2014).

Recall that premise 1 of the evolutionary argument against naturalism is the claim that the probability of the reliability of human cognition on the conjunction of naturalism and evolution is low. In other words, if both naturalism and evolution were true, argues Plantinga, one should expect human cognition to be unreliable. And it is *this* – if true – that he thinks will furnish the naturalist with an ultimately undefeatable epistemic defeater for her belief in naturalism.[4] The rest of the argument shows how this might work, and it runs (roughly) as follows:

If the naturalist's cognitive faculties are unreliable, then *any* belief she forms will also be unreliable. For *every* belief she forms would ultimately be the product of her belief-forming faculties – by definition. And since naturalism and evolution would be among those beliefs, they will likewise be unreliable. And since *every* belief is, or would be, unreliable, there appears to be no way in which the proposed defeater *itself* could be defeated. For to defeat *it*, the relevant epistemic agent would need to employ the very faculties whose epistemic credentials have been undermined in the first place. In other words, if *every* belief – including defeaters – ultimately require reliable cognitive faculties to underwrite its warrant or justification, then every belief which *isn't* the product of such faculties wouldn't be justified. In essence, it is ultimately the *universality* of the cognitive unreliability involved which confers the particularly pernicious epistemic defeater Plantinga thinks the naturalist will acquire. This, in a nutshell, is the evolutionary argument against naturalism.

2. Plan of the thesis

As noted, the thesis is split into two main parts. In Part 1, the truth or falsity of premise 1 – also known as the 'probability thesis' – is the going concern. In Part 2, I ask which of evolutionary naturalism or theistic evolution best explains the evidence we have with respect to the reliability of human cognition. As noted above, I think that the second question is important and relevant to this work in that if it can be shown that evolutionary naturalism is indeed the better explanation of the relevant empirical evidence, then evolution, far from being naturalism's kryptonite, is in fact a very close friend. Part 1 occupies chapters one and two, and part 2 chapters three and four. The work concludes in Chapter 5.

3. Part 1: The probability thesis

In Chapter 1, Plantinga's argument in support of premise 1 will be discussed. As briefly highlighted, his argument draws its inspiration from the idea that evolution is primarily interested in a creature's survival, not the truth-value of their beliefs. And although common sense strongly suggests that *what* one believes is crucial to *how* one behaves, Plantinga claims that this is unlikely to be the case on naturalism and evolution. For he argues that there

[4] In general, an epistemic defeater is 'a condition' which leads to a belief losing 'positive epistemic status', suffering a reduction in such status, or never acquiring such a privileged epistemic position in the first place. Epistemic defeaters come in two basic forms; *rebutting* defeaters and *undermining* defeaters. Rebutting defeaters are defeaters which would give one reason or reasons to withhold, or forego, belief in *p* – indeed, they would give one positive reasons to believe *not-p*. Undermining defeaters – which are the type of defeaters Plantinga appeals to in the EAAN – gives one reason or reason to doubt the *grounds* for believing *p*. In other words, it gives one a reason or reasons to think that the reasons one has for believing *p* aren't any good (Sudduth, 2020).

are no naturalistically-friendly scenarios – bar one – on which belief – *as belief* – enters the causal chain leading to behaviour. And on that scenario on which it does, there is no good reason to think that evolution would mostly select for true belief. In short, Plantinga argues that if belief content doesn't enter the causal chain leading to behaviour as *content*, then evolution cannot select for it, and hence cannot select for its truth-value either. And where it *can* select for belief content, its unlikely to select for *true* content more often than not.

If Plantinga is right, then it would be a serendipitous piece of good fortune if adaptive behaviour and true belief were as closely linked as a miser is to his money. And if this holds true on *every* naturalistic scenario available to the naturalist in accounting for the relation between belief and behaviour, then it would be true on *all* of them. Hence, it seems that human cognition would indeed be unreliable on naturalism and evolution. Premise one of the evolutionary argument against naturalism would be true.

Chapter 2 consists in a detailed presentation, discussion, and evaluation of a number of naturalist responses to the probability thesis. Some have argued that premise 1 is either false, empty, or has no truth-value. Others acknowledge that evolution's primary care may be for creatures' survival and not the truth-value of their beliefs, but argue that what best *explains* the differential adaptivity of different behaviours are the truth-values of the beliefs that inform them. Yes, behaviour may be what *ultimately* matters, but they maintain that *what* one believes will be crucial in determining what one *does*, and hence be (indirectly) exposed to evolution's selective pressure. Thus, according to them, evolution *does* care about truth, even if not ultimately.

4. Part 2: Which is the better explanation of human cognitive reliability, evolutionary naturalism or theistic evolution?

In Chapter 3, I present an overview of the scientific evidence with respect to the reliability of human cognition. I also entertain a number of proposed evolutionary naturalist explanations thereof. As you will see, human sensory-perception is not the reliable register of nature's affairs that one may have initially supposed it to be. Neither is human reasoning without its epistemic maladies; people often and systematically fall prey to cognitive biases of many kinds, including reasoning in ways that depart from what the norms of rationality suggest they should. Finally – and most importantly in terms of this thesis – there is evidence that suggests that individuals aren't epistemically reliable when it comes to subject matters far removed from those relevant to their survival – e.g. metaphysics.

The evolutionary-naturalistic proposals I consider all take as their explanatory point of departure the assumption that the human mind should be seen as evolution's answer(s) to natures recurring adaptive challenges. In short, say these theorists, one can learn a great deal about how the human mind works – including making sense of evidence to be presented here – if one thinks of human cognition as an evolutionary adaptation. As you will see, such an 'adaptationist' approach has much to recommend it. But in light of this thesis, what I find most compelling is the evolutionary naturalist's explanation of humans' evident lack of reliability with respect to "godly" subject matters. For, as I will argue, I think it is *here*, if nowhere else, where the contrast in explanatory power between evolutionary naturalism and theistic evolution is most apparent.

In Chapter 4, I turn to a number of theistic responses to the problems the naturalist thinks the empirical evidence discussed in Chapter 3 raises. Specifically, I argue that the theist who believes that God guided the process of evolution shouldn't find it unduly troubling that humans don't reason in optimally rational ways or that their sensory-perceptual faculties are not as reliable as they probably could have been. An all-powerful and perfectly good being – such as Plantinga believes God to be – may have good reasons for creating us as less than optimally rational, reasons we are not privy to at present. On the other hand, I argue that there is as yet no plausible theistic response to the significant problem that humans evidently don't know much about God, if anything at all.

In Chapter 5 I close. With respect to Part 1 of the thesis, I conclude that premise 1 is false; human cognition would likely be reliable on naturalism and evolution. And the objection that I think manages to defeat it trades on the idea that *even* if evolution cannot select for belief content, there are plausible conceptual reasons to think that true belief and adaptive behaviour would likely be favourably related.

To the question posed in Part 2 of this work, I answer that it may in fact be that theistic evolution is as good an explanation of the evidence discussed as any naturalistic alternative, perhaps even better on some counts. But to my mind, the fact that it misses on the knowledge-of-God data point ultimately makes it the lesser of the two.

CHAPTER 1: THE PROBABILITY THESIS

1. Introduction

Premise 1 of the evolutionary argument against naturalism is the claim that human cognition would likely be unreliable if naturalism and evolution were true – i.e. that one wouldn't be able to trust that human cognitive faculties produce mostly true beliefs on naturalism and evolution. In this chapter, premise 1 of the evolutionary argument against naturalism (EAAN) will be discussed. Plantinga's aim is to show that human cognitive reliability would likely be unreliable on (metaphysical) naturalism and (contemporary) evolutionary theory (Plantinga, 2011b). Before considering Plantinga's arguments in support of premise 1, we first need to consider its nature – its form, and what it consists in. This will be the subject of discussion in Section 2.

The intuition that leads Plantinga to formulate premise 1 as he does – what he refers to as 'Darwin's doubt' – will be the subject matter of Section 3 (Plantinga, 2011b: 316, 317). In short, Darwin's doubt is the idea that evolution primarily cares about the *behaviour*, and not the *beliefs*, of creatures navigating their respective environments. According to Plantinga, this raises the possibility that true belief and behaviour, or even just belief and behaviour, whatever its truth-value, are not as straightforwardly linked as common sense suggests.

In sections 4 and 5, Plantinga considers various naturalistic proposals concerning the putative relations that hold between belief and behaviour, and what their effect on the probability of human cognitive reliability would likely be.[5] According to Plantinga, three possibilities exhaust the naturalist's response with regards to the nature of these relations: epiphenomenalism, reductive physicalism, and non-reductive physicalism (Plantinga, 1993: 223, 224; Plantinga, 1994: 6, 7; Plantinga 2011b: 322). Plantinga argues that there's no reason to think natural selection is able to select for greater cognitive reliability on any of these proposals – hence, according to him, there's no reason to think human cognitive faculties would be reliable on evolutionary naturalism (Plantinga, 1993: 223, 224; Plantinga, 2011a: 442, 444).

2. The form and content of premise 1

Premise 1 of the evolutionary argument against naturalism is the claim that the probability (P) of the reliability (R) of human cognitive faculties on the conjunction of metaphysical naturalism (N) and contemporary evolutionary (E) theory is low. In symbols, P(R | N&E) is low. This is a probability claim – i.e. a claim of a

[5] Plantinga is aware that there are complexities attending to the matter of belief:

> A materialist might hold that belief-talk is to be paraphrased into talk about the property of believing; then we could say that human beings sometimes display the property of believing *p,* for some proposition *p,* while denying that there are any such things as *beliefs.* For what follows, this difference makes no difference. Eliminativism is also an option for the physicalist. In this paper, though, I will be assuming that there really are such things as beliefs (or at any rate ways of believing), because what we are investigating is an argument for the claim that a certain belief (or way of believing), namely the belief that N&E, is rationally unacceptable (Plantinga, 2011a: 436)

probabilistic nature. More specifically, it's a conditional probability statement. Further, it's "substance" involves the notions of human cognitive reliability (R), metaphysical naturalism (N), and contemporary evolutionary theory (E).

2.1 The form of premise 1

A conditional probability is the probability that some event A will occur given that some other event B *has* occurred. In mathematical terms, this is written as P(A | B). Premise 1 of the EAAN is of this general form – being P(R | N&E) – the only difference is that the state conditioned on – the state of affairs that needs to obtain – is a conjunction, N *and* E, as opposed to a singular event or state of affairs B (as in P(A | B)). This is a difference that makes no difference – premise 1 is a prime example of a conditional probability claim.

2.2 The concepts employed in formulating premise 1

2.2.1 Human Cognitive Reliability

According to Plantinga, the reliability of a creature's cognitive faculties (R) is a function of the number of true beliefs vis-à-vis false ones that a creature holds to, believes, or comes to believe. Plantinga also states that a given creature has reliable cognitive faculties if the ratio of the number of true-to-false beliefs it entertains equals or exceeds 2-to-1 (Plantinga, 2011b: 313).[6] Elsewhere he is less specific, suggesting that cognitive faculties are reliable 'if the great bulk of its deliverances are true' (Plantinga, 1994: 3). In general, the thought seems to be that human cognitive faculties are reliable when they produce true beliefs at a significantly greater rate than chance – i.e. 1-to-1 or 50/50. And cognitive faculties are:

> [T]hose faculties, or powers, or processes, that produce beliefs in us. Among these faculties (are) memory... perception... apriori intuition... sympathy... introspection... testimony... (and) induction .... [M]any would add that there is a moral sense (as well) (Plantinga, 2011: 311, 312).

2.2.2 Metaphysical Naturalism

The nature of naturalism – what it *is* – has been the subject of extensive discussion (Papineau, 2020). *Metaphysical* naturalism is that part of naturalism concerned with what reality fundamentally consists in – what its ontological ingredients are (Papineau, 2020). As yet, there is no agreed upon definition – no set of necessary and sufficient conditions picking out metaphysical naturalism (and it only). However, there is general agreement that no Gods, ghosts, or anything ontologically similar are allowed. Plantinga says as much:

> [I]t isn't easy to say what precisely naturalism *is*, but perhaps this isn't necessary in this context (i.e. of

[6] Why he suggests that cognitive faculties are reliable when they produce true beliefs at a rate of 2-to-1 isn't clearly explained, but it makes no difference to the substance of the discussion that follows.

> the evolutionary argument against naturalism). Crucial to metaphysical naturalism, of course, is the view that there is no such person as the God of traditional theism (or anyone like him) (Plantinga, 1993: 220, my emphasis).

2.2.3 Contemporary Evolutionary Theory

In broad outlines, sufficient for purposes of the evolutionary argument against naturalism, the contemporary theory of evolution involves two notions: common descent and natural selection. Common descent is the claim that all biota ultimately originated from a single, or small number, of primordial replicators (Theobald, 2012). Natural selection is generally taken to be the primary mechanism that explains common descent, although additional mechanisms such as genetic drift, sexual selection, and neutral evolution, amongst others, have been proposed (Theobald, 2012). Natural selection operates by means of *random* genetic mutations, which either prove fitness enhancing, neutral, or deleterious with respect to the creature in which such mutations occur or has occurred (Loewe & Hill, 2010).[7] Importantly, Plantinga is adamant that the *random* genetic mutations involved in the theory of evolution – as *science* – does not entail that the process of natural selection is *not* 'being caused, orchestrated (or) arranged by God' (Plantinga, 1994: 3). For evolution guided by God – and hence involving a much *weaker* sense of randomness than pure chance – would be empirically indistinguishable from evolution *not* so guided by him or anyone like him. As he puts it:

> [The contemporary theory of evolution is] entirely compatible with the thought that God has *guided* and orchestrated the course of evolution, planned and directed it, in such a way as to achieve the ends he intends. Perhaps he causes the right mutations to arise at the right time; perhaps he preserves certain populations from extinction; perhaps he is active in many other ways (Plantinga, 2011b: 308, his emphasis).

Hence, according to Plantinga, whether evolution is guided (weakly random), or unguided (truly random), is a philosophical add-on to evolutionary science. It's *not* an indispensable part of the *science*.

3. Darwin's doubt

In Darwin, Plantinga finds support for the idea that evolution is primarily concerned not with the reliability of creatures' cognitive faculties, but with their *behaviour.* Hence, the moniker 'Darwin's doubt' (Plantinga, 2011b: 316, 317). Darwin's worry concerning human cognitive reliability – given their genealogy – is clear:

---

[7] Note that the "randomness" referred to here is not that of the pure chance, as Loewe and Hill explain:

> It has often been said that mutations are random, a statement that is simultaneously true and false: true because mutations do not originate in any way or at any time that is related to whether their effects are beneficial – one of the central tenets of Neo-Darwinism; and false because mutations are the result of complex biochemical reactions that result in non-uniformly distributed mutation frequencies, favouring some (random) changes over others (Loewe & Hill, 2010: 1154).

> With me, the horrid doubt always arises whether the convictions of man's mind, which has been developed from the mind of the lower animals, are of any value or at all trustworthy. Would any one trust in the convictions of a monkey's mind, if there were any convictions in such a mind (Darwin, 1881)?

Patricia Churchland claims much the same to be the case – that evolution is primarily in the business of fitness enhancing behaviour, not in generating true belief – when she states that:

> Boiled down to essentials, a nervous system enables the organism to succeed in the four F's: feeding, fleeing, fighting, and reproducing. The principal chore of a nervous system is to get the body parts where they should be in order that the organism may survive… [I]mprovements in sensorimotor control confer an evolutionary advantage: a fancier style of representing is advantageous *as long as it is geared to the organism's way of life and enhances the organism's chances of survival*. Truth, whatever that is, definitely takes the hindmost (Churchland, 1987: 548, 549, her emphasis).

In the EAAN, Plantinga employs Darwin's doubt to set the stage for premise 1; *if* evolution is primarily interested in creature's behaviour, and not in the truth-value of their beliefs, then the question whether human cognition is likely trustworthy on evolutionary naturalism appears to attain critical mass (Plantinga, 2011b: 316, 317). More specifically, Plantinga thinks that Darwin's doubt gives one reason to wonder whether belief and behaviour are as tightly packaged or causally related as everyday intuition seems to suggest.

In section 4, Plantinga suggests how this might work. He claims that there are number of scenarios or possibilities of how belief and behaviour may be related that differs from the everyday or 'common sense' view.[8] Moreover, he argues that these scenarios are more probable than one might initially suppose.

4. Scenarios with respect to the relation or relations between belief and behaviour

Plantinga thinks the proposals naturalists have to offer with respect to the nature of the possible relation or relations that hold between belief and behaviour fall into three (broad) categories: epiphenomenalism, reductive physicalism, and non-reductive physicalism (Plantinga, 1993: 223, 224; Plantinga, 1994: 6, 7; Plantinga, 2011b: 322).[9]

[8] Where the 'common sense' view is roughly the idea that beliefs and desires cause our actions, and that true beliefs are on average better guides to successful behaviour – achieving our goals or fulfilling our desires – than false ones.

[9] In Plantinga (2011) he notes that:

> Materialists offer fundamentally *two* theories about the relation between physical and mental properties... reductive materialism and nonreductive materialism (Plantinga, 2011b: 322, my emphasis).

However, in Plantinga (1993) it's clear that he thinks that epiphenomenalism is also a possible position that the naturalist who thinks there are such things as beliefs may also adopt in explaining the relation between belief and behaviour.

4.1 Epiphenomenalism *simpliciter* and semantic epiphenomenalism

Epiphenomenalism is the idea – in the philosophy of mind – that beliefs don't have any causal potency. More specifically, beliefs don't enter causal chains as a result of their *content*. If they do feature in such chains, it's purely as the *effects* of causally sufficient *physically* (instantiated) properties (presumably neurophysiological properties in humans).[10] On epiphenomenalism, the relation between belief and the causally efficacious physical properties is permissive in one sense, and stringent in another. On the one hand, belief is independent of the causally efficacious (physical) goings-on, and, according to Plantinga, can be about *anything* (Plantinga, 2011b: 331). On the other hand, epiphenomenalism strictly forbids the *content* of belief, *as content*, to affect causal change.

For example, consider Bob. Suppose Bob wants a beer. On epiphenomenalism, belief and desire won't be things that *result* in either Bob's desire for beer or his getting up from the couch and getting himself a beer, given that he believes there's beer in the fridge and the best way of getting said beer is making his way to the fridge – i.e. getting off the couch and so on. Belief, *as belief*, and desire, *as desire*, won't be the things that *cause* him to be successful or not in getting what he wants given what he believes (Plantinga, 1993: 223).[11]

In summary; Bob's beer drinking might not be a prime example of fitness enhancing behaviour, but the general point should be clear; behaviour, on epiphenomenalism, whether fitness enhancing or not, is caused by physical properties, but beliefs, *qua* beliefs, are causally impotent. Further, on epiphenomenalism, beliefs could either be completely independent of any (physical) causal chains leading to – or from – behaviour, or they could be independent as causal factors, but be side *effects* of such chains (Plantinga,1993: 223, 224). Belief would either be like a type of mental aether, undisturbed by the physical, and not disturbing anything in turn, or be like water bubbles on the surface of boiling water, the effect of whatever physical is happening below, but not feeding back into this physical set-up as a cause.

Within epiphenomenalism, Plantinga distinguishes between what he calls epiphenomenalism *simpliciter* and *semantic* epiphenomenalism. On epiphenomenalism *simpliciter*, beliefs don't cause anything – they are causally irrelevant. On semantic epiphenomenalism, beliefs *do* enter the causal chain leading to behaviour *as causes*, but only on account of their syntax (physical properties), not their semantics (content or meaning) (Plantinga, 1993: 224). The syntactical properties of a belief will be implemented as neurophysiological properties of some sort – e.g. structures or events comprising complex arrays of neurons, their firing rates, the brain chemistry that mediates the transfer of neural signals, and so on (Plantinga, 1994: 7, 8). The semantic properties would be the property of

[10] Epiphenomenalism can be understood as a type of *property* dualism; on this view, there is a clear difference between *mental* properties and *physical* properties, but it remains a *substance* monism in that only a physical substance is taken to exist (Plantinga, 2011a: 436).

[11] Mental content properties (such as belief) are things that can be picked out or identified by phrases such as the belief *that* p for some proposition *p*, in this case Bob's belief that there is beer in the fridge.

having various beliefs (or desires) – e.g. the belief that *p*, where *p* is some proposition.

For example, given epiphenomenalism *simpliciter*, Bob would be successful in his beer-getting-behaviour without so much as a whiff of the involvement of the mental – the only requirement would be his having physical features sufficient to the task. On semantic epiphenomenalism, Bob's successful behaviour would be the result of his belief's physical or syntactic properties only, not their semantics. Moreover, on both epiphenomenalism *simpliciter* and its semantic cousin, the content of Bob's belief could have been about *anything* (Plantinga, 2011b: 331). For all his beer-getting behavioural success, he could have entertained beliefs about anything whatsoever – e.g. perhaps the location of Jimmy Hoffa's body or the location of Atlantis (Plantinga, 1993: 223; Plantinga, 2011b: 331).

Briefly; Plantinga's claim is that on neither of epiphenomenalism *simpliciter* nor semantic epiphenomenalism is the link between belief content and behaviour causal. On neither does belief *cause* behaviour. As a result, he argues that that natural selection would have no purchasing power on the reliability of creatures beliefs. From this he concludes that there would be no reason to think that creatures beliefs will be reliable on evolutionary naturalism and epiphenomenalism (Plantinga, 2011b: 330).

4.2 Reductive physicalism (RP)

Reductive physicalism is the view that beliefs are identical to, or reducible to, physical properties.[12] In the case of human beings, these would presumably be neurophysiological properties of some sort (Plantinga 2011b: 324). In Bob's case, this amounts to the claim that whatever the physical make-up of his cognitive faculties, the beliefs that he entertains at any given moment – e.g. that there is beer in the fridge – will be *identical* to some neurophysiological property of his. In short, Bob's mental properties just 'are, or are reducible to, neurophysiological properties' (Plantinga, 2011b: 324). The physical and the mental are not different *things*, but the *same* thing described in different *words*.[13]

According to Plantinga, there are two (general) ways in which one could make sense of the relation between belief and behaviour consistent with reductive physicalism. One, beliefs, being identical to physical properties, could be causally efficacious, but result in *maladaptive* behaviour on the part of creatures that come to hold and act on such

[12] There is no single agreed upon definition on what reductive physicalism is Stoljar (2017). The view closest to the one Plantinga endorses appears to be what's referred to as type physicalism. Type physicalism is the view that:

> For every actually instantiated mental property *F*, there is some physical property *G* such that *F* = *G* (Stoljar, 2017).

Whatever the exact differences may be – if any – Plantinga's definition will be the one employed in this work.

[13] More accurately, on reductive physicalism, the mental and physical is one thing (substance), but there is a property such it is *both* the property of having such-and-such content *and* the property of being the physical property or properties identified with such-and-such content (Plantinga 2011a: 436).

beliefs (Plantinga, 1994: 7, 8). This can happen in two ways:

(a) The causally efficacious beliefs may be unnecessary, and thus be an evolutionary expensive addition to the creatures' cognitive establishment. It would have been better for them not to have them (Plantinga, 1994: 8).

(b) The relevant beliefs may prove deleterious, but not excessively so (Plantinga, 1993: 223). For example, a group of humans – monks of the self-flagellating kind perhaps – may hold the belief that the occasional application of the whip to their sinful flesh would increase the odds of them enjoying a favourable outcome in the hereafter. However, they may remain oblivious to the fact that the costs of this peculiar activity outweigh the benefits (assuming it does). Still, it's possible that this activity has no serious fitness-reducing consequences.

Two, the everyday or common sense view may be true – i.e. beliefs may be causally efficacious *and* prove fitness enhancing (for the most part at least) (Plantinga, 1994: 8). Surprisingly, Plantinga thinks that the probability of human cognitive faculties being reliable on this scenario is much lower than one may initially suppose (Plantinga, 1994: 8). How so?

He notes that it's not *only* belief that leads to behaviour, but also desire.[14] Plantinga claims there are a great number of belief-desire pairs where the belief is false but the desire is such that it results in the requisite fitness enhancing behaviour (Plantinga, 1994: 8). In other words, if one adds an appropriate desire to a false belief (a false-belief-appropriate-desire pair), all will end well for the creature concerned, at *least* as well as if the creature acted on a true-belief-appropriate-desire pair. Plantinga introduces Paul the hominid to show how this might work:

> Perhaps Paul very much *likes* the idea of being eaten, but whenever he sees a tiger, always runs off looking for a better prospect because he thinks it unlikely that the tiger he sees will eat him. This will get his body parts in the right place so far as survival is concerned, without involving much by way of true belief. (Of course we must postulate other changes in Paul's ways of reasoning, including how he changes belief in response to experience, to maintain coherence.) Or perhaps he thinks the tiger is a large, friendly, cuddly pussycat and wants to pet it; but he also believes the best way to pet it is to run away from it… (Plantinga, 1994: 8, his emphasis).

According to Plantinga, insofar as it is a possibility that false beliefs can be combined with the right desires to result in adaptive behaviour – indistinguishable from behaviour that would result from true beliefs and desires – the probability that the common sense view obtains is commensurably diminished (Plantinga, 1993b: 226). Hence, it's not clear, as many would initially suppose, that causally efficacious belief leading to adaptive behaviour implies that those beliefs leading to those adaptive behaviours are *true* (for the most part). Hence, the probability of the reliability of human cognitive faculties on common sense should be reduced accordingly.

[14] More may be involved in intentional behaviour than the conjunction of belief and desire or beliefs and desires. For present purposes, the important point is that it's not *only* belief that informs behaviour.

It may be the case that the naturalist neither holds to epiphenomenalism nor reductive physicalism, either for the reasons Plantings raises, or for other reasons commonly raised against these views in the literature.[15] As a result, many naturalists are non-reductive physicalists; holding that the relation between belief and the behaviour is one of supervenience.

4.3 Non-reductive physicalism (NRP)

Non-reductive physicalism is the view that belief content *supervenes* on physical properties. This means that what beliefs are about, *cannot* (by either causal, metaphysical, or logical necessity) change without there been a commensurate change in the *sub*vening physical properties (Plantinga, 2011b: 323).[16] And if a belief *does* change, there *must* be some sort of change in the subvening physical properties. Moreover, any number of different physical properties or states can give rise to a particular belief, in a many-to-one relationship. Conversely, on non-reductive physicalism, it's not possible for *many* beliefs to be realized given any *singular* physical state.

For example, if Bob were in a world where non-reductive physicalism was true, his original belief – that there is beer in the fridge – *cannot* be about something else, that there is a cat on the mat for example, given the *same* physical properties that did the relevant causal work. If the belief is a different one, the physical properties *must* be different. However, it could be that Bob has a different physical constitution, alien perhaps, and hence different physical properties in his new alien "brain", but still have exactly the same mental property – the belief that there is beer in the fridge.

According to Plantinga (2011b: 330), there are two possibilities – consistent with non-reductive physicalism – in making sense of the relation between the beliefs of humans, or any creatures relevantly similar in cognitive

[15] For example, a well-known objection to reductive physicalism is referred to as the argument from multiple realizability. Multiple realizability is the idea that it's possible that *different* physical structures (alien cognitive neurophysiology perhaps) may give rise to the *same* mental properties. Hence, at first glance, it appears untrue that a given physical state – e.g. a human neurophysiological state or states – can be *identified* with some belief or beliefs (Bickle, 2020).

[16] More formally, and in general, there are three types of supervenience relation and they are defined as follows:

> (WS) Weak Supervenience: M weakly supervenes on P just in case necessarily for any object *x* and any property F in M, if *x* has F, then there exists a property G in P such that *x* has G, and if any *y* has G, it has F.
>
> (SS) Strong Supervenience: M strongly supervenes on P just in case necessarily for any object *x* and any property F in M, if *x* has F, then there exists a property G in P such that *x* has G, and necessarily if any *y* has G, it has F.
>
> (GS) Global Supervenience: M globally supervenes on P just in case for any two worlds, w1 and w2, if they are P-property indistinguishable, then they are M-property indistinguishable (Ritchie, 2008: 117).

Clearly, this means that other types of property, not only the mental, can supervene on still others, hence supervenience is not exclusively a concept used in the philosophy of mind. For example, it's also often employed in ethics to characterize the relationship that holds between ethical and physical properties

endowment, and the physical properties on which they supervene. One, it may be that their beliefs don't enter the causal chain leading to their behaviour *at all*. On this scenario, the truth-value of their beliefs would be epiphenomenal in virtue of being unrelated, or independent, of the complete set of physical states or processes relevant to these creatures' behaviour. Even though the relevant physical properties will *determine* these creatures beliefs, 'natural selection just has to take potluck with respect to the (resultant) propositions or content determined by those adaptive neurophysiological properties' (Plantinga, 2011b: 330, his emphasis). The reason being that the content determined is 'just be a matter of logic or causal law, and natural selection can't modify either' (Plantinga, 2011b: 330).

Two, it could be that belief contents are merely the *effects* of causal chains leading from external and/or internal environmental stimuli to neurophysiological states, but are themselves causally mute (Plantinga, 1994: 6). It could also happen that the physical properties (neurophysiological properties for example) – and the beliefs that are correlated with them – are the result of a *common* physical cause. Either way, Plantinga maintains that there is no reason to suppose that the belief content so correlated would be *true* (Plantinga, 1994: 7). As he puts it:

> The *neurology* causes adaptive behaviour and also causes or determines belief content: but there is no reason to suppose the belief *content* thus determined to be true (Plantinga, 2011b: 327, my emphasis).

On either of these scenarios, the causal efficacy of belief content – *as content* – is irrelevant to the evolutionary fitness of these creatures. Natural selection is singularly efficacious with respect to *physical* properties only (Plantinga, 2011b: 331). For example, consider Bob the beer lover again. He has the belief that there is beer in the fridge and he wants it. He does the necessary to get it. On the twin possibilities canvassed above, the *physical* properties do all the causal work. The belief content associated with these physical properties is causally impotent.

Recall that Plantinga thinks that the belief content associated with the neurophysiology (of humans or creatures relevantly similar) doesn't have to be true, it doesn't even have to be '*about* the objects involved in the states of affairs causing the subvening properties' (Plantinga, 2011b: 331, his emphasis). For example, in Bob's case, this would amount to claiming that he could be equally successful in his beer-getting behaviour even if the content of his beliefs isn't about beer *at all*.

The discussion thus far has been focussed on the possible scenarios Plantinga thinks the naturalist has recourse to in explaining the nature of the relation or relations between belief and behaviour (semantic epiphenomenalism, reductive physicalism, and non-reductive physicalism).[17] But, as yet, nothing has been said about the probability of the reliability of human cognitive faculties on each of these scenarios. Further, the question regarding the probability that each of these scenarios are true (on naturalism and evolution) hasn't been addressed.

[17] This is not strictly true; these aren't the *only* possibilities open to the naturalist in accounting for the link between belief and behaviour (see note 16 for details).

5. Human cognitive reliability on semantic epiphenomenalism, reductive physicalism, and non-reductive physicalism

5.1 Introduction to disaggregation

At this point, the reader should have a clear grasp of the elements of the probability thesis. But what is the probability of the reliability of human cognitive faculties on semantic epiphenomenalism, reductive physicalism, and non-reductive physicalism? And further, what is the probability that each of these scenarios is true on naturalism and evolution? Finally, how do the twin questions raised above relate to the question of what the probability of reliability of human cognition on naturalism and evolution is?

The details of Plantinga's argument in answering the questions posed above are best seen in relief. By disaggregating the probability thesis into its component parts a framework is created within which Plantinga's arguments with respect to the probability of the reliability of human cognition on each of the different scenarios, and the probabilities of each scenario on naturalism and evolution, is rendered clearly.

By the theorem of total probability, P(R | N&E) can be disaggregated into three separate terms, each of which is one of the scenarios naturalists have to offer in accounting for the nature of the relation between belief and behaviour. Expressed mathematically:

$$P(R \mid N\&E) = [P(R \mid SE \text{ and } N\&E) \times P(SE \mid N\&E)] + [P(R \mid RP \text{ and } N\&E) \times P(RP \mid N\&E)] + [P(R \mid NRP \text{ and } N\&E) \times P(NRP \mid N\&E)]$$[18]

In English, this equation states that:

The probability of the reliability of human cognitive faculties on naturalism and evolution – P(R|N&E) – the term on the left hand side of the equation, is equal to the sum of three terms (those on the right hand side of the equation).

---

[18] In its simplest and most correct (disaggregated) form, premise 1 of the evolutionary argument against naturalism (EAAN) should have only two terms on the right; a term for views on which belief content is causally efficacious, and one's on which they're not (Plantinga, 2002a: 9, 10). The terms C and not-C represent these two mutually exclusive and jointly exhaustive possibilities:

$$P(N \mid E) = P(R \mid C) \times P(C \mid N\&E) + P(R \mid \text{not-}C) \times P(\text{not-}C \mid N\&E)$$

The three views discussed in the text – i.e. semantic epiphenomenalism, non-reductive physicalism, and reductive physicalism – aren't the *only* possibilities available to the naturalist in accounting for the relation between belief and behaviour, and are thus not jointly exhaustive. For instance, it doesn't include epiphenomenalism *simpliciter*.

Moreover, as far as I know, there are no responses to the evolutionary argument against naturalism that don't appeal to some form of non-reductive physicalism or reductive physicalism in objecting to premise 1 of Plantinga's argument. Or at least no responses that assume that there are such things as beliefs.

The first term on the right is the conjunction of P(R | semantic epiphenomenalism and N&E) and P(semantic epiphenomenalism | N&E). The first conjunct of this first term on the right – P(R | epiphenomenalism and N&E) – represents the contribution that semantic epiphenomenalism would make to the probability of the reliability of human cognitive faculties if it were true. The second conjunct of the first term – P(semantic epiphenomenalism | N&E) – represents the probability that semantic epiphenomenalism is the relation that holds between belief and behaviour on naturalism and evolution.

The second term on the right is the conjunction of P(R | reductive physicalism and N&E) and P(reductive physicalism | N&E). The first conjunct of this second term – P(R | reductive physicalism and N&E) – accounts for the contribution that reductive physicalism would make to the probability of the reliability of human cognitive if it were true. The second conjunct of this term – P(reductive physicalism | N&E) – represents the probability that reductive physicalism would hold on naturalism and evolution.

The final term on the right hand side of the equation is the conjunction of P(R | non-reductive physicalism and N&E) and P(non-reductive physicalism | N&E). The first conjunct of this final term on the right – P(R | non-reductive physicalism and N&E) – represents the contribution that non-reductive physicalism would make to the probability of the reliability of human cognition. The second conjunct of the final term – P(non-reductive physicalism | N&E) – represents the probability that non-reductive physicalism would be true on naturalism and evolution.

Given this disaggregation of premise 1, it is now a relatively simple matter to inspect the arguments Plantinga gives in support of his ultimate claim – that P(R | N&E) is low. One only needs to discuss the arguments made in support of each of the three terms represented on the right hand side of the equation expressed above.

5.2 The probability of the reliability of human cognition on semantic epiphenomenalism

Should the naturalist be concerned about the probability of the reliability of human cognition on semantic epiphenomenalism? Specifically, what is the probability that semantic epiphenomenalism is true given naturalism and evolution? And what is the probability that human cognition would be trustworthy on semantic epiphenomenalism, naturalism, and evolution? Concerning the latter, Plantinga thinks that it's low:

> ... [I]f semantic epiphenomenalism is true, it will not be the case that a false belief causes maladaptive behaviour by virtue of its having false content, and it will not be the case that a true belief causes adaptive behaviour by virtue of having true content. The truth or falsehood of belief will then be irrelevant to fitness and thus, so to speak, *invisible* to natural selection (Plantinga, 2011a: 437, his emphasis).

Conversely, Plantinga claims that the probability of semantic epiphenomenalism on naturalism and evolution is high, as:

> [I]t is exceedingly difficult to see… how they (beliefs) can enter that (causal chain) *by virtue of their*

> *content: a given belief it seems, would have had the same causal impact on behaviour if it had had the same* (physical) *properties, but different content* (Plantinga, 2011a: 436, his emphasis).

Plantinga offers analogies in support of these claims (Plantinga, 2002b: 214, 218). He asks us to imagine an opera singer breaking a crystal glass as a result of her singing a series of high notes. He suggests that it's easy to see how this might happen as a result of the *physical* properties of her voice, but that what she is singing *about* appears to be causally irrelevant (Plantinga, 2002b: 214). Or, he suggests, imagine that bricks only came in white. According to him, when any particular brick breaks a window, it wouldn't in virtue of its colour, but its physical properties only – i.e. momentum and such (Plantinga, 2002b: 218).

If Plantinga is correct – that the probability of the reliability of human cognition *is* low on naturalism and evolution, *and* that the probability of semantic epiphenomenalism is *high* on the same, P(R | N&E) would be low, *irrespective* of the probability values of the second and third terms in the disaggregated equation (Plantinga, 2011a: 437).[19] The reason is that if the first term on the right colonizes a sufficiently large part of the available probability space, then there wouldn't be enough room for the second and third terms to make a difference to the probability of the reliability of human cognition. Specifically, if semantic epiphenomenalism is sufficiently likely on naturalism and evolution, then it wouldn't matter if human cognition is very likely on reductive or non-reductive physicalism. For they – reductive and non-reductive physicalism – wouldn't be likely *enough* to save human cognition from probable unreliability. In other words, if semantic epiphenomenalism is likely true on naturalism and evolution, and if the probability of the reliability of human cognition is low on semantic epiphenomenalism, then the probability thesis would be true – the probability of the reliability of human cognition on naturalism and evolution would be low.

But suppose that Plantinga is mistaken with regards to semantic epiphenomenalism, such that P(R | N&E& semantic epiphenomenalism) isn't low or that P(semantic epiphenomenalism | N&E) isn't high. This would mean that the values of P(the negation of semantic epiphenomenalism | N&E) and P(R | N&E& and the negation of semantic epiphenomenalism) will then become relevant. In short, P(reductive physicalism | N&E), P(R | N&E& reductive physicalism), P(non-reductive physicalism | N&E), and P(R | N&E& non-reductive physicalism), will have to be accounted for.

5.3 The probability of the reliability of human cognition on reductive physicalism

Recall that on reductive physicalism, mental content properties enter the causal chain leading to behaviour due to the identity of the physical and the mental.[20] More specifically, on reductive physicalism, there would be a property such that it is *both* a physical property *and* the property of being the belief that *p*, where *p* is a proposition.

[19] Naturally, the probability of the second term and third term has to cohere with the axioms of the probability calculus, which means that it has to be sufficiently *low* to ensure the mathematics remains consistent.

[20] Note that *properties* – as abstract objects – aren't causally efficacious *as* properties, only their physical representatives. For example, it's brain states – neural structures and firing patterns – that enter causal chains, not the properties (mental and physical) which they *represent*.

This property is causally efficacious as a result of its physical properties, but given that it is a property that is *both* physical *and* the belief that *p*, the belief will also enter the causal chain leading to behaviour.

The identity of the mental and the physical appears to secure the idea that evolution *has* purchase on the belief content of cognitively sophisticated creatures. This invites one to think that evolution would weed out creatures that entertain evolutionary wayward thoughts vis-à-vis their comrades that don't. Plantinga demurs:

> [A]ssume that having P (a neurophysiological property identical with having a belief with content *p*) is adaptive in that it helps cause adaptive behaviour. But (given no more than N&E& reductive physicalism), we have no reason at all to suppose that this content, the proposition *p* such that P is the property *having p as content*, is *true*; it might equally well be false. True… the property *having p as content* does indeed enter the causal chain leading to behaviour; but it doesn't matter, as far as adaptiveness goes, whether this first bit of content is true. What matters is only that the neurophysiological property in question causes adaptive behaviour; whether the content it constitutes is also *true* is simply irrelevant. It can do its job of causing adaptive behaviour just as well as if it is false as if it is true; it doesn't matter (Plantinga, 2011a: 444, his emphasis).

And:

> Naturalism, evolution, and reductive physicalism (N&E&RP) gives us no *connection* between the *truth-value* of the content and the *adaptiveness* of the behaviour it causes… (Plantinga, 2011a: 441, 442, my emphasis).

In other words, what Plantinga is saying here is that even were one to metaphysically identify the content of belief with its neurophysiology, there wouldn't be any good reason to expect the content so identified to be true. For even though some belief will of necessity be identical to a given naturally-selected-neurophysiology, the survival of the relevant creature will be a function of the latter only.

Put differently, Plantinga is claiming that a creature's reproductive success is only a matter of it having the right sort of neurophysiology. For he argues that the identity relation available to the reductive physicalist in explaining the link between belief and behaviour only tells us that belief content properties will of necessity share an identity relation with neurophysiological properties. It gives us no reason why *this* belief is identical with *this* neurophysiology or *that* belief with *that* neurophysiology. Hence, if Plantinga is right, there would seem to be no reason on account of this identity relation between belief and behaviour to expect true beliefs to be connected to adaptive neurophysiology more often than not. In fact, Plantinga concludes – employing the principle of indifference – that it's just as likely that the content associated with the underlying neurophysiology in this way would be true or false (Plantinga, 2011b: 334).[21]

[21] The principle of indifference... states that in the absence of any relevant evidence, a rational agent will distribute their credence (or 'degrees of belief') equally amongst all the possible outcomes under consideration (Benjamin, 2019). Concerning this principle, Fitelson and Sober claim that:

Assuming that the average person holds to at least two independent beliefs, the probability that human cognition would be reliable would be low, and become vanishingly unlikely as they entertain more beliefs (Plantinga, 2011b: 335). Hence, if reductive physicalism were true, or likely true, the probability of the reliability of human cognition on naturalism and evolution would be low or very low.

5.4 The probability of the reliability of human cognition on non-reductive physicalism

Recall that on non-reductive physicalism, mental content properties supervene on physical properties by causal, metaphysical, or logical necessity. In the case of human beings, this would mean that the beliefs they entertain would likely supervene on their neurophysiology. Further, on non-reductive physicalism, there cannot be a change in any of their beliefs without some change in their neurophysiology. How does the nature of this relation inform the question with respect to what the probability of the reliability of human cognition on nonreductive physicalism and evolutionary naturalism is? Moreover, what is the probability that non-reductive physicalism is true on naturalism and evolution?

Plantinga acknowledges that evolution can mould behaviour such that it proves adaptive, and that belief content will accrue to such adaptive neurophysiology consistent with the supervenience relation. However, he harbours serious doubts that evolution would be able to channel the belief content that so supervenes in the direction of

---

> Bayesians have never been able to make sense of the idea that prior probabilities have an *objective* basis. The siren song of the Principle of Indifference has tempted many to think that hypotheses can be assigned probabilities without the need of empirical evidence, but *no consistent version of this principle has ever been articulated*. The alternative to which Bayesians typically retreat is to construe probabilities as indicating an agent's subjective degree of belief. The problem with this approach is that it deprives prior probabilities (and the posterior probabilities that depend on them) of probative force. If one agent assigns similar prior probabilities – (that the probability of the truth value of belief content identified with adaptive neurophysiology is .5 for example) - ,(it) is entirely consistent with another agent's assigning very unequal probabilities to them, if probabilities merely reflect intensities of belief (Fitelson & Sober, 1998: 116, my emphasis).

In other words, if Plantinga claims that that it is *objectively* – that everyone considering the issue *must* agree - that the probability that the truth-value of belief content identified with adaptive neurophysiology is .5, he confronts the challenge that 'no consistent version of this principle has ever been articulated'(Fitelson & Sober, 1998: 116, my emphasis). And if he retreats to viewing probability claims as 'reflections of intensities of belief', he faces the objection that different assignments of prior probabilities is consistent with someone else assigning a different prior (probability) to this (his) proposition (Fitelson & Sober, 1998: 116, my emphasis).

Plantinga replies that:

> The Bertrand paradoxes show that certain incautious statements of the principle of indifference come to grief – just as Goodman's grue/bleen paradoxes show that incautious statements of the principle governing the projection of predicates or properties come to grief. Still, the fact is we project properties all the time, and do so perfectly sensibly. And the fact is we also regularly employ a principle of indifference in ordinary reasoning, and do so quite properly. We also use it in science – for example in statistical mechanics (Plantinga, 2011b: 332).

greater reliability. As he puts it:

> This new (mental content) property will be implied with causal or metaphysical necessity by the relevant neurophysiological property which, we may assume, is adaptive; but that doesn't give us a ghost of a reason for assuming that the content thus accruing to the structure is *true*... Natural selection is obliged to take potluck; it selects for adaptive neurophysiological properties, but must then accept the content properties, true or false as the case may be, that supervene on them. Non-reductive physicalism doesn't specify or imply *any connection between content and adaptivity*, and indeed no natural connection comes to mind (Plantinga, 2011a: 444, my emphasis).

And:

> Natural selection can modify the neurophysiological properties in the direction of greater fitness, but that doesn't mean or make probable that the consequent modification of the supervening content properties is towards truth (Plantinga, 2011a: 445).

In other words, Plantinga is claiming that evolution can select for adaptive neurophysiology, but not for content. And hence not for true content either. It has to accept whatever content inevitable comes "superveniently" attached to the neurophysiological selections it makes. Moreover, there is nothing on non-reductive physicalism *itself* – on the supervenience relation *sans* evolution – that suggests that belief content and *adaptive* behaviour would likely be related systematically. And hence no reason to think that *true* content would likely be associated with adaptive behaviour more often than not.

Plantinga concludes from the above that that one should therefore expect the probability of the reliability of human cognition on non-reductive physicalism and evolutionary naturalism to be low or very low (Plantinga, 2011b: 331). Further, if the reliability of human cognition *is* low on evolutionary naturalism and non-reductive physicalism, it wouldn't matter what the probability of non-reductive physicalism on naturalism and evolution is, the conjunction of P(R | N&E and non-reductive physicalism) and P(non-reductive physicalism | N&E) would be low.

5.5 The probability of semantic epiphenomenalism, reductive physicalism, and non-reductive physicalism on naturalism and evolution

The disaggregation of premise 1 – see page 25 – indicates that one also needs to consider how probable each of semantic epiphenomenalism, reductive physicalism, and non-reductive physicalism is (on naturalism and evolution). The expected reliability of human cognition on any of these scenarios is a function both of how reliable human cognition is likely to be (on naturalism, evolution, and that specific scenario), *and* how likely that scenario is on naturalism and evolution.

As discussed, Plantinga claims that one shouldn't expect human cognition to be reliable on *any* of semantic

epiphenomenalism, reductive physicalism, or non-reductive physicalism (each conjoined with naturalism and evolution). If Plantinga is right, then it wouldn't matter how probable or improbable any of the respective scenarios are on naturalism and evolution. For, given that they are mutually exclusive, jointly exhaustive, and subject to the dictates of probability theory, the probability of the reliability of human cognition on naturalism and evolution would be low.[22]

6. Conclusion

The discussion in Chapter 1 centred on premise 1 of the evolutionary argument against naturalism – that the probability of the reliability of human cognition on metaphysical naturalism and evolution is low. According to a number of naturalists, including Darwin, evolution is primarily interested in the behaviour of creatures, not the reliability of their beliefs or belief forming faculties. Plantinga calls this Darwin's doubt. If evolution only cares about behaviour, and behaviour is fully accounted for in terms of the physical (on naturalism and evolution), where does that leave belief, belief which common sense strongly suggests *is* causally efficacious?

According to Plantinga, the naturalist has three options open to her in aiming to explain the nature of the relation between belief and behaviour – semantic epiphenomenalism, reductive physicalism, and non-reductive physicalism. However, Plantinga argues that there's no reason to suppose or expect that the beliefs related to the physical in these three possible ways is, or will mostly be, true. In fact, he argues that one should estimate the probability that humans have reliable cognition on each scenario to be .5 at most. Assuming that these scenarios exhaust the naturalists' responses in accounting for the relation between belief and behaviour, the implication follows that their aggregation – this aggregation being the probability that human cognitive faculties are reliable on naturalism and evolution – will also be low. Hence, premise 1 of the evolutionary argument against naturalism would be true – i.e. the probability of the reliability of human cognition on naturalism and evolution would be low. A number of naturalist challenges to Plantinga's argument in support of premise 1 follow.

[22] More accurately, the three views discussed are not quite exhaustive. For, as noted – c.f. footnote 16 – epiphenomenalism *simpliciter* is also an option available to the naturalist in explaining the link between belief and behaviour. Thus, strictly speaking, the probability of the reliability of human cognition on epiphenomenalism *simpliciter* should also be considered. However, according to Plantinga, the arguments that should lead one to expect human cognition to be unreliable on semantic epiphenomenalism would apply – *mutatis mutandis* and *salva veritate* – to epiphenomenalism *simpliciter* as well (Plantinga, 2002).

If this is right, the distinction between semantic epiphenomenalism and epiphenomenalism *simpliciter* wouldn't make a difference in determining whether premise 1 of the evolutionary argument against naturalism is true or not. For if the same probability arguments hold with respect to both views, then, at least for purposes of calculating the relevant probabilities, they would be indistinguishable. I think Plantinga made the distinction to show that even if some "part" of a belief *does* enter the causal chain leading to behaviour, there still wouldn't be any good reason to think that the link between the content of a belief and its causal structure would be one congenial to human cognitive reliability.

CHAPTER 2: RESPONSES TO PREMISE 1 OF THE EVOLUTIONARY ARGUMENT AGAINST NATURALISM

1. Introduction

In this chapter, the more prominent arguments naturalists have offered in response to premise 1 of the evolutionary argument against naturalism will be presented and explored. A number of naturalists have argued that it is false – i.e. that the probability of the reliability of human cognition on naturalism and evolution *isn't* low, but sufficiently high to secure its trustworthiness.

Before delving into the arguments that are raised against the probability thesis *as it stands*, consideration will be given to those arguments that take issue with *how* it stands. Respondents have objected to Plantinga's framing of each of the concepts employed in the formulation of the probability thesis. More specifically, they claim that Plantinga's characterization of human cognitive reliability is overly simplistic, his distinction between 'guided' and 'unguided' evolution is confused, and his conception of naturalism empty (Boudry & Vlerick, 2014; Childers, 2011; Van Fraassen, 2006). The arguments in support of these claims are discussed in Section 2.

The arguments for the falsity of premise 1 claim that natural selection, *contra* Plantinga, *is* causally effective in shaping belief in general, and sufficiently effective in funnelling human cognition in the direction of greater reliability specifically. Moreover, a number of respondents have claimed that even supposing the link between belief and behaviour *non*-causal, there are nonetheless plausible reasons for thinking the probability of the reliability of human cognition sufficiently high on evolutionary naturalism. Arguments in support of these claims are the subject matter of Section 3.

However, to secure this conclusion – that human cognition is likely trustworthy on naturalism and evolution – naturalists also need to show that the scenarios on which human cognition is expected to be reliable are themselves likely on naturalism and evolution. Arguments focussed on meeting this challenge are the focus of section 4

2. Objections to the formulation of premise 1

A number of naturalists have taken issue with the formulation of the probability thesis, some concluding that the argument 'fails to launch' as it stands, others that the framing of premise 1 is overly simplistic. Moreover, the distinction Plantinga makes between 'guided' and 'unguided' evolution has been challenged as misguided.

2.1 Is there no question of naturalism?

The problem of defining metaphysical naturalism is familiar.[23] It's the challenge of picking out metaphysical naturalism, and only it, among everything else there is – including its competitors and compatriots. Despite its

[23] When I use the term naturalism, I refer to metaphysical naturalism unless otherwise specified.

familiarity and vintage, the lively debate it continues to engender indicates that it retains a significant degree of philosophical interest.[24] The literature aimed at giving naturalism its due is vast, and certainly beyond the scope of this thesis. However, naturalism is clearly a significant part of Plantinga's project – a project against *naturalism* no less – and hence he owes the critic a reasonable response if challenged.

Here I will consider Bas Van Fraassen's – and indirectly Alyssa Ney's – objections to Plantinga's argument. Theirs are examples of arguments to the effect that naturalism or physicalism is best construed not as a metaphysical *thesis*, but as an *attitude* or cluster of *attitudes* (to science) (Van Fraassen, 1996; Ney, 2008).[25] In short, their argument is that an "attitudinal" view allows the naturalist or physicalist to sidestep Hempel's well-known dilemma – a dilemma facing anyone who wishes to take naturalism as a thesis about what is or isn't the case. Hempel's dilemma runs as follows: If the naturalist takes the 'physical' to be what contemporary fundamental physics says it is, then her naturalism is false, as the twin theories at the heart of contemporary fundamental physics – quantum physics and general relativity – cannot both be right (Wilson, 2006: 65). On the other hand, if she thinks the 'physical' is whatever a future completed physics says it is, then her position will either be 'obscure', or perhaps include entities that a flag-bearing naturalist would consider fundamentally non-physical (Witmer, 2012: 103, 104). As Plantinga thinks that 'metaphysical naturalism... is (at least) the view that there is no such person as the God of traditional theism', his would be an example of the latter (Plantinga, 1993: 220).

In what follows, Hempel's dilemma will be presented and explored in the light of a number of arguments aimed at its dissolution. Objections to these proposed "dissolutions" will also be considered, with particular attention being paid to what I will refer to as the 'attitudinalist' or 'attitudinal' views of Ney and Van Fraassen. The importance of the attitudinalist view with respect to this work lies in the fact that it has been employed as a direct challenge to Plantinga's evolutionary argument. The objections Van Fraassen's raises with respect to Plantinga's naturalism, and Plantinga's response thereto, will draw the discussion in Section 2.1 to a close.

2.1.1 Taking the bull by the horns: responding to Hempel's dilemma

A well-known response aimed at the first horn of Hempel's dilemma – appealing to contemporary physics to get a fix on the physical – is that of Andrew Melnyk (2003). He argues that physicalism is essentially a scientific hypothesis, which he thinks allows the physicalist, like the scientific realist, to remain rational in holding to physicalism (or scientific realism), even though she can acknowledge that her view is likely false (Witmer, 2012: 104). Witmer responds that it's not clear that Melnyk's response 'resolves (Hempel's) dilemma' (Witmer, 2012:

[24] See Papineau (2020).

[25] I use the term naturalism and physicalism interchangeably in this thesis. However, the "naturalist dualism" of David Chalmers illustrates that they may come apart (Witmer, 2012: 90; Chalmers, 1996). In support of his status as a naturalist Chalmers argues that the fundamental phenomenal entities he suggests exist are – like physical entities – 'governed by laws of nature and... subject to similar causal explanations' (Chalmers, 1996; Witmer, 2012: 99). Having said that, their identification in this work makes no difference to the arguments considered herein.

104). For as Witmer notes; '*both* the scientific realist and the would-be physicalist face a problem: shouldn't they be able to tell us what they *do* think is likely true (Witmer, 2012: 104)?

Van Fraassen raises another sort of objection to those – like Melnyk – who appeal to contemporary physics to define the physical. It runs as follows:

> ... [S]upposing the empirical claim(s) (of the contemporary physical theory appealed to)... is empirically investigated and is found wanting or false, will there or will there not be a fall-back position to call the real materialism after all? It would be a poor game if after much strife, the loser could say 'that's not it at all, that's not what I meant at all... is that the end of materialism?... (No)... a favourite belief of the materialists (physicalists) would have to be relinquished, but they would all know how to retrench. For the *spirit* of materialism is never exhausted in piecemeal empirical claims (Van Fraassen, 1996: 166, 167, my emphasis).

In other words, Van Fraassen is claiming that the physicalist arguing in this manner may *claim* that their physicalism is a hypothesis comparable to those of science, but he is doubtful that they would forego their physicalism if that "hypothesis" proves false. I don't think this is much of an argument, but rather a claim that physicalists who maintain that their physicalism is a falsifiable hypothesis are only pretending. When the evidence shows their hypothesis to be untenable, 'they would all know how to retrench', instead of admitting that their physicalism has been defeated. This may or may not prove to be the case. Still, it doesn't undermine the claim that one could formulate one's physicalism in this manner. For the *content* and *formulation* of the relevant arguments should demand the critic's attention, not the *motives* of those who make them.

Other responses to Hempel's dilemma have either aimed to blunt the second horn of the dilemma, or tried to outflank both. Examples of the first stipulate that 'the "physical" is *not* to include anything mental unless it is nothing over and above something else that qualifies as physical' (Witmer, 2012: 104). This strategy would rule out any unwanted fundamentally mental entities from the start, thus precluding the problem that a future ideal physics might deal in fundamentally (mental) entities from being a problem (Witmer, 2012: 104). Further, as Witmer notes, this (stipulative) move would also help 'somewhat with the obscurity objection, since one can then know something about what is *not* included in the physical' (Witmer, 2012: 104, his emphasis).

A line of argument similar to the above – called the 'Via Negativa' – is another important response aimed at tackling the second horn of the dilemma (Montero & Papineau, 2005; Witmer, 2012; Stoljar, 2017). In short, it defines the physical in an entirely negative way – i.e. in terms of what it is *not*. Specifically, it cashes out the physical in terms of the non-mental; 'everything is either non-mental or nothing over and above the non-mental (Witmer, 2012: 103, 104). In a more formal vein, Stoljar thinks that the Via Negativa can be seen as stating the definition of 'the notion of a physical property; something like: *F* is a physical property if and only if *F* is a non-mental property' (Stoljar, 2017).

However, he notes that there are a number of reasons to avoid such a definition. The reason is that that there are properties that are neither mental nor physical – e.g. the *élan vital* of vitalism (Stoljar, 2017). This creates a

problem as there might have been a world in which animals and plants instantiated this property, and still, so he claims, 'one should not say on this account that plants and animals instantiate a mental (or a physical property) – i.e. the élan vital is neither mental nor physical' (Stoljar, 2017). In other words, if the *élan vital* is not a physical property, then on the above definition, it *must* be a mental property. But it's not clear that were plants and animals to instantiate this non-physical property, that they *would* be instantiating a mental property (Stoljar, 2017). If Stoljar is right, the Via Negativa as stated cannot accommodate this fact and needs to be revised (Stoljar, 2017).

He suggests that 'one might try to meet this objection by revising the Via Negativa so that what is intended is only a partial definition' (Stoljar, 2017). Something like: *F* is a physical property only if *F* is non-mental (Stoljar, 2017). But, according to him, this would create problems of its own. For, as he argues, there might be properties that are both mental *and* physical (Stoljar, 2017). Indeed, if the type-identity theory in the philosophy of mind is true, every mental property – such as a wish or a belief – would be identical to some physical property. But this wouldn't be allowed on the partial definition considered – there *couldn't* be a property such that it is both a mental and physical property. How so? Suppose there were such a property or properties. On the partial definition this would lead to: *F* is a physical property only if *F* is non-mental *and* mental (Stoljar, 2017). Or, put differently, because some properties are both mental *and* physical, the partial definition would lead to: F is a physical-and-mental property only if F is *not* a mental-and-physical property. But this can't be (Stoljar, 2017).

In responding to those who appeal to a future ideal science to characterize the physical, Van Fraassen says the following:

> If you press a materialist, you quickly find that the most important constraint on the meaning of the Thesis – i.e. physicalism – is that it should be compatible with science, *whatever science comes up with...* (But), he certainly does not know *what* he believes. For of course he has no more idea than you or I of what physics will postulate in the future. It is a truly courageous faith that believes in 'I know not what' - isn't it? Indeed, in believing this, (the physicalist) cannot be certain that he believes anything at all. Suppose science goes on forever, and every theory is eventually succeeded by a better one. That has certainly been the case so far, and always some accepted successor has implied that the previously postulated entities... do not exist. If that is how it will continue, world without end, then (of course, there wouldn't be an ideal completed physics, and no question of what such a theory would postulate)... (Van Fraassen, 1996: 167, 168, his emphasis).

Van Fraassen's argument here is the same as those who argue that appealing to a future ideal physics will result in a claim that is obscure. For he maintains – like many others – that one wouldn't know what such a physics will 'postulate in the future'. And hence, evaluating the truth or falsity of a physicalism wedded to an unknown ideal physics would be a fruitless exercise.[26] Witmer claims much the same when he notes that:

[26] Further, Van Fraassen claims that it may *never* be possible for the truth or falsity of such a physicalism to be determined. For he argues that every physical theory thus far proposed has been replaced by another. And what is to say that this pattern of continual replacement will not prove to be indefinite? Or that we would be able to discover or invent the ideal theory even if there is one?

> If we don't know what the ideal physical theory looks like, we don't know what sorts of properties will count as physical properties; how, then, can anyone support, attack, or otherwise assess the physicalist thesis (Witmer, 2012: 104).

Above, I have given a brief overview of the sorts of arguments and objections that have been raised with respect to Hempel's dilemma. Suppose that none of the arguments that take naturalism as a thesis are successful – i.e. that none can be said to offer a plausible exit strategy in the face of Hempel's dilemma. What other plausible options, if any, are available to the physicalist?

2.1.2 The attitudinal view

Given the seemingly robust challenge facing anyone who wishes to escape the horns of Hempel's dilemma, a number of philosophers have argued that naturalism should be construed not as a thesis about what *is* or *isn't* the case, but as an *attitude* or 'cluster of attitudes' (Ney, 2008; Van Fraassen, 1996). They argue that taking naturalism or physicalism as an attitude allows one adopting such an attitude to remain undisturbed by Hempel's concerns. For an attitude cannot be said to be true or false, and hence, wouldn't be troubled by either horn of Hempel's dilemma. As Alyssa Ney's explains:

> Although it – i.e. (adopting the right attitude towards physics) – does understand 'physics' as current physics, it avoids both horns since by giving up status as a doctrine, the view cannot be false or trivial; attitudes are not properly evaluable as true, false, or trivial (Ney, 2008: 10).

But what, might one ask, does this attitude amount to? Ney answers:

> (One ought to take) physicalism (as an) attitude one takes to form (one's) ontology completely and solely according to what physics says exists. It is a commitment to... swear to go in (one's) ontology everywhere and only where physics leads (one) (Ney, 2008: 9).

In short, staking one's claim to metaphysical naturalism as a doctrine would, according to Ney, not be as effective as viewing it as an attitude, for an attitude cannot be true or false, and hence Hempel's dilemma will not trouble the naturalist. Further, in the spirit of naturalism, the "right" attitude for a naturalist to assume should be one of deference to science in matters of ontology. Where science goes, the naturalist – *qua* naturalist – should follow.

Moreover, Ney claims that – even were Papineau *et al* successful in giving a definition or characterization of the physical – it would be a historical regression, and thus, all else equal, be inferior to viewing it as an attitude. As she puts it:

> There is something peculiar about Papineau's proposal, namely the clear sense in which it involves a historical regression... Physicalism took over the reins from materialism because philosophers wanted a positive ontological theory despite the fact that they could not be sure what kind of entities physicists would in the end posit in their theories. Therefore, it was best to hand over the ontological authority to the physicists and not enforce any *a priori* constraints on what type of entities they may sensibly say the

> world contains. By removing the link between physics and physicalism, we are taking a step backward... (Ney, 2008: 9, my emphasis).

In other words, naturalists are in large part *naturalists* because they think science has been more successful than philosophy in telling us what reality consists in. Hence their claim that philosophy shouldn't tell science how the world *must* be, but wait on science to show how it *actually* is. By defining the physical such that science is precluded from ever considering mental entities to be fundamental, Papineau-and-company appear to violate this self-imposed naturalist credo.

In what follows, Bas Van Fraassen's attitudinal view – and Plantinga's response thereto – will be discussed. In *specie*, Van Fraassen's attitudinal view is no different from Ney's. Both think that Hempel's dilemma can be successfully avoided only if the naturalist conceives of her naturalism as an attitude or cluster of attitudes to science, as opposed to a thesis about what is or isn't the case. As noted, the importance of including Van Fraassen's view in this work is that he argues for it in the context of opposing *Plantinga's* naturalism – a version of naturalism that is of a kind with those that make claims about what is or isn't the case. Further, Van Fraassen's argument and Plantinga's specific response to the former's objections has the virtue of reducing the risk that I might misrepresent the latter's argument were I to respond on his behalf.

2.1.3 Bas Van Fraassen's objections to Plantinga's Evolutionary Argument against Naturalism

It is in light of the difficulties Hempel's dilemma confers on those who think naturalism is a metaphysical view that Bas Van Fraassen's objections to Plantinga's evolutionary argument against naturalism should be understood. Van Fraassen argues that Plantinga's evolutionary argument against naturalism 'fails to launch', as naturalism isn't a view about what *is* or *isn't* the case, but an *attitude* to science (Van Fraassen, 2006: 170). More specifically, he claims that '… no such argument as Plantinga has recently given against naturalism can succeed' (Van Fraassen, 1996: 172). The reason being that naturalism 'is not identifiable with a theory about what there *is*, but only with an *attitude*, or cluster of attitudes' (Van Fraassen, 1996: 170, my emphasis). Hence, to ask whether naturalism is true or false wouldn't make sense, with the downstream result that asking whether human cognition should be expected to be reliable on naturalism wouldn't admit of a meaningful answer either (Van Fraassen, 1996: 170, 172).

Further, he argues that Plantinga's claim that naturalism is – if nothing else – the view that there's no such person as God or anyone like him cannot be 'taken as a definition, on pain of the circularity involved in characterizing (the) "natural" in terms of (the) "supernatural"'(Van Fraassen, 1996: 172). Plantinga responds by claiming that:

> Perhaps there are no explicitly stateable and reasonably precise necessary and sufficient conditions for something's being material (natural), but there may be necessary conditions (Plantinga, 2006: 350).

In other words, naturalism may not be *clearly* identifiable, such that it is *exactly* such-and-such and *only* such-and-such, but that doesn't mean it admits of *anything,* or *nothing*. Just as traditional theism clearly precludes God from being a hot air balloon or a spaghetti monster – whatever or whomever God may otherwise be – naturalism

is *at least* the view that the central claim of traditional theism is false – i.e. that no omnipotent, omniscient, omnibenevolent, and omnipresent being usually referred to as 'God' exists (Plantinga, 2006: 350).[27]

I think that Plantinga is right. Even if we don't have precise truth-coordinates where we might find it, it doesn't mean that we have no clue as to where it might be, or know where it can't be. For all we know, the keys to naturalism may be hidden almost anywhere, but it's very unlikely that it will be found in the house of God.

2.2 Plantinga's alleged simplistic framing of the reliability of human cognition

Another set of respondents have claimed that it's overly simplistic to ask whether the probability of the reliability of human cognition is high or low, that is, that the answer to the question posed by Plantinga can be a simple yes or no (Boudry & Vlerick, 2014; Childers, 2011). According to them, to treat this question – as to the probability of the reliability of human cognition on naturalism and evolution – fairly, requires a far more nuanced, and empirically supported, approach (Boudry & Vlerick, 2014: 66). As they put it:

> Plantinga forces us to make a stark choice between two mutually exclusive propositions, to wit R ('our cognitive faculties are reliable') or its negation, not-R ('our cognitive faculties are not reliable') (Boudry & Vlerick, 2014: 66).

And Childers:

> Such cartoonish scepticism arises only as a by-product of forcing ourselves to think in artificially rigid black and white categories (whether human cognition is reliable or not *tout court*), while ignoring our best current naturalistic models of the biological and social evolution of cognition (Childers, 2011: 201, my emphasis).

What these claims aim to highlight is that such a seemingly 'cartoonish' framing – that human cognition is either reliable or not *across the board* – is to neglect scientific findings that have offered answers of a much more fine-grained nature. Answers such as *which* cognitive faculties or processes are less than reliable, and *when*, *where*, *why,* and about *what* one should find or expect them to be so.

For example, in contexts in which humans take mind-altering drugs, the mind *is* altered, which often results in them forming strange (and presumably false) beliefs about what is or isn't the case in their immediate environment. And where one introduces unaided or untrained human cognitive faculties to contexts far removed from those in which they evolved – such as the performances of a skilled conjurer – one can expect that the beliefs humans form as a result will likely be unreliable. It's unlikely that humans *won't* believe what their senses are telling them (that it appears that the Statue of Liberty has disappeared for example). But it's also unlikely that they *will* believe what they are seeing. Perhaps it's this tension between seeing-believing and believing-what-

[27] By traditional Christian theists at least, which is the kind of theism referred to throughout this work.

you're-seeing that makes a "show" of magic so appealing to many.

In short, naturalists raising this objection are claiming that Plantinga is ignoring what to them is of primary importance, *scientific* evidence, evidence which they claim show that the question as to the probability of the reliability of human cognition admits of no simple unqualified answers (Boudry & Vlerick, 2014: 66; Childers, 2011: 201).

However, if human cognitive faculties *are* probably unreliable on naturalism and evolution, then appealing to empirical evidence to determine in which contexts they can be expected to be reliable appears confused. For if these faculties are unreliable, *any* beliefs they generate will likewise be infected with unreliability, or so Plantinga claims (Plantinga, 2002: 11, 12). As a result, it would appear that the naturalist would find herself in epistemic quicksand were her aim to divorce science from the charge that the probability of human cognition on evolutionary naturalism is low. For science itself would be a product of those very faculties whose reliability is under suspicion – i.e. science would be no less a product of human cognitive faculties than any other beliefs human beings may entertain, and hence would fall under the same cloud of suspicion as they.

Simply put, if the expected value of human cognitive reliability is low, it wouldn't matter how that reliability or unreliability is *distributed* among the contexts in which human cognitive faculties are expected to function. It wouldn't change the fact that human cognition is probably unreliable. Moreover, and as a result, one wouldn't be able to trust the deliverances of science with respect to the distribution of reliability vis-à-vis unreliability anyway, even if one thought it important. Hence, naturalists raising this objection – employing science to raise the expected reliability of human cognition or objecting to the 'simplistic' framing of the quandary – would be raising an objection that is irrelevant at best, or begging the question at worst. In effect, naturalists would be assuming science reliable when it is the result of the very faculties whose reliability has been undermined. Or they would be highlighting in which contexts human cognition is, and can be expected to be, more or less reliable, a claim with which Plantinga can readily agree (Plantinga, 1993: 18).

2.2.1 The unreliability of native human cognition vis-à-vis the reliability of augmented human cognition

Boudry and Vlerick claim that aiming to establish the epistemic credibility of evolutionary naturalism given the unreliability of human cognition is neither question begging nor irrelevant (Boudry & Vlerick, 2014: 73). According to them, Plantinga is too hasty in moving from the premises that 'evolution won't produce organisms that produce (overall) true belief' and 'evolutionary naturalism is itself a product of those cognitive faculties' to the conclusion that 'evolutionary naturalism is therefore self-defeating' (Boudry & Vlerick, 2014: 73). They argue that:

> Underlying this argument is the assumption, which we call the 'foundationalist fallacy', that if the foundations of our cognitive endeavours are not *completely* secure, then the whole edifice built on top of it must collapse. If the ground is even a *little* shaky, no amount of scaffolding will be sufficient (Boudry & Vlerick, 2014: 73, their emphasis).

In response to what they take to be Plantinga's 'foundationalist fallacy', Boudry and Vlerick argue – following John Clendinnen (1989) – that one's 'epistemic confidence' (in some proposition) can be gradually raised as evidence accumulates' (Boudry & Vlerick, 2014: 73). How so? Consider propositions P, Q, and R. Suppose these propositions are related such that 'if P then Q, if Q then R, and if R then P'. Then they point out that:

> In a deductive model of justification, it is often incorrectly assumed that each step is either completely justified or not justified at all. However, it is possible to have an interdependence of justifications (P, Q, R) in which we start out by an initially very weak and provisional version of P, whereby our confidence is gradually raised through Q and R (Boudry & Vlerick, 2014: 73).

What Boudry and Vlerick are claiming here is that one doesn't have to be absolutely *sure* that *each* proposition (or any) one employs in an argument is 'completely' justified. For they argue that the epistemic credentials an individual belief has can be raised by it being related to others in an appropriate way. The justification of an individual proposition can be supported or improved by others with which it stands in a certain logical relation or relations.[28]In other words, theirs is a type of holism, where the parts of an argument can support one another to the epistemic benefit of the whole. They apply this line of reasoning to the question of the reliability of human cognition when they argue that:

> … [B]y starting from the deliverances of our bare and unassisted mind, including folk psychology and basic perceptual capacities, we may gradually increase the reliability of our beliefs, by gathering evidence… (Boudry & Vlerick, 2014: 73).

Thus, even if Plantinga were correct – that unguided evolution probably endowed human beings with unreliable cognition – this would only apply to the *unassisted* working of such faculties. But this doesn't mean that humans having such faculties cannot happen upon tools and practices that assist them in their survival. With time, such rudimentary practices and tools may *eventually* grow more sophisticated as those who use them (individually, or more likely, as a cognitive community) have to continue their more or less random search for new answers to natures ever-changing questions (Childers, 2011: 201). Childers explains how this might happen:

> As our knowledge advances, so do our methods of acquiring knowledge. Nobody comes into this world with the idea of the double-blind experiment, but nearly anyone can *learn* it. Technological innovations lead to progress in the sophistication of our empirical-theoretical models, which leads to further technological growth. This advancement of knowledge stems not from having reliable cognitive faculties, but from being good at imitation and learning by trial and error. As our species learns more, more becomes learnable. Because genetics and innate brain physiology alone fail to determine our beliefs, we cannot answer questions about the reliability of our *individual* cognitive faculties without

[28] It's also reasonable to think that the justification a given proposition has can be *reduced* in a similar fashion.

> taking into account the quality of our *collective* cultural knowledge... (Childers, 2011: 20, his emphasis).

Boudry and Vlerick make a similar point. The theory of evolution is not the product of an individual's unassisted mind, but the result of the cumulative and intergenerational efforts of many learning from, and feeding off, each other's mistakes and insights. As they put it:

> The theory of evolution by natural selection is by no means the product of our bare, unassisted mind… It is supported by a massive body of knowledge accumulated over many generations… aided by formal scientific systems of analysis… and technological innovations… (Boudry & Vlerick, 2014: 73).

In other words*, if* the theory of evolution were the product solely of 'unaided human cognition', then expecting it to be unjustified would be correct. However, evolutionary theory is the product of processes informed by the scientific enterprise – itself a communal product of trial and error learning, technological progress, and cultural transmission. If humans were able to raise their epistemic reliability in the manner Childers, Boudry, and Vlerick suggests, the result would likely be a cognitively reliable community. The community will "know", even if the individual (largely) doesn't.

However, can one be confident that – even if the theory of evolution is the product of 'aided' human cognition – that the 'aids' that presumably bootstrap it to epistemic respectability are in fact successful, or can be expected to be so? It's certainly the case that science is a very impressive epistemic enterprise, furnishing humans with an ever increasing body of knowledge, that, if perhaps not true with a capital T, has allowed them an unparalleled mastery of nature.

But, although humans have been imminently successful in augmenting their native perceptual faculties with all sorts of technical devices, it's not clear that they *would* have been able to do so in a world where evolutionary naturalism were true. For, one cannot just assume that *this* world is place where evolutionary naturalism *is* true – this needs to be argued for as it's not clear that it is. As Plantinga puts it:

> We are not asking about how things *are,* but about *what things would be like if both evolution and naturalism... were true*. We are asking about P(R | N&E), not about P (R | the way things actually are). Like everyone else, I believe that our cognitive faculties are for the most part reliable, and that true beliefs are more likely to issue in successful action than false one's. But that's not the question. The question is what things would be like if N&E were true; and... we can't just assume, that if N&E were true, then things would still be the way the way they are (Plantinga, 2011b: 335, 336, his emphasis).

Further, and perhaps a thornier issue, is the question whether individual human cognition *would* be sufficiently reliable to get the epistemic bootstrapping process going in the first place, as Boudry and Vlerick supposes (Boudry & Vlerick, 2014: 73). Specifically, Boudry and Vlerick assume that 'we may gradually increase the reliability of our beliefs, by gathering evidence' (Boudry & Vlerick, 2014: 73). And that on the basis of inferential schemas like 'if P then Q, if Q then R, and if R then P'...'it's possible to have an interdependence of justifications (P, Q, R)', whereby one's confidence in P may be raised (Boudry & Vlerick, 2014: 73). Still, can one just assume

this to be the case? For wouldn't it be question begging to take it that inferential procedures like 'if P then Q, if Q then R, and if R then P' *are* reliable? It appears so. For it would be assuming that the very inferential procedures, practices, and cognitive faculties needed to discover, invent, or employ P, Q, and R, *are* reliable, when *that's* what one needs to be shown.

Having said that, this would only be a cutting or fatal objection were (these) naturalists trying to convince the *global* sceptic – that perpetual and unredeemable cynic of the epistemic world.[29] Moreover, premise one is the claim that human cognition is probably unreliable, not that it is *completely* unreliable – i.e. Plantinga doesn't claim that human cognition is completely or utterly depraved, only *likely* so (Plantinga, 2011b: 314). And herein lies the necessary room for the naturalists' to escape the charge of (vicious) circularity, and hence allow their arguments to remain live options for those who want to oppose Plantinga's epistemically dismal conclusion. Hence, if the (relevant) naturalists are correct, individual unreliability wouldn't imply, without further argument, that the *community* of "unreliables" is similarly fated to suffer epistemic misery.

If the charge of circularity doesn't stick, these arguments appear forceful, for the assumptions on which they are founded don't appear unreasonable. What is required is the (sufficiently) reliable transmission of (minimally) reliable beliefs, and a world in which those who form, hold to, and act on true beliefs are differentially and adequately rewarded for so believing and acting compared to those who don't (cf. Vlerick 2012; Boudry & Vlerick, 2014: 68; Ramsey, 2002: 19). On evolutionary naturalism, these requirements appear unproblematic. The onus falls on Plantinga to show that the reliability of human knowledge – those of the individual, but especially the collective – cannot tend to greater reliability.

On the naturalist arguments considered above, humans are able to achieve collective knowledge, – at least with respect to science – despite their (individual) cognition being unreliable in the manner Plantinga claims. The states of affairs or conditions required for coming to communal knowledge – despite individual unreliability – are three-fold.

One, the relevant prehistoric cognitive agents should be able to form, hold to, and act on at least *some* minimum number or ratio of true beliefs – i.e. the degree of the unreliability of their belief-forming faculties and the beliefs they generate should not be too excessive; some (minimum) level of reliability is required. Two, reliable beliefs should confer an adaptive advantage on those who form, hold to, and act on true beliefs vis-à-vis those who don't.

[29] As Vlerick (2012) argues, *every* epistemology is ultimately circular. Everyone *has* to start with some claim to knowledge in order to get their epistemological enterprise off the ground – 'a theory of knowledge cannot start from nothing' (Vlerick, 2012: 181). As Bertrand Russell puts it:

> If we adopt the attitude of the complete sceptic, placing ourselves wholly outside all knowledge, and asking, from this outside position, to be compelled to return within the circle of knowledge, we are demanding what is impossible, and our scepticism can never be refuted. For all refutation must begin with some piece of knowledge which the disputants share; from blank doubt, no argument can begin. Hence the criticism of knowledge which philosophy employs must not be of this destructive kind, if any result is to be achieved. Against this absolute scepticism, no logical argument can be advanced (Russell, 1912: 112).

Three, the relevant agents should be sufficiently capable and successful in adopting each other's beliefs and practices and be able to transfer these beliefs and practices with enough fidelity to their progeny or the progeny of their conspecifics. If these conditions hold, humans may trust that their collective cognitive efforts are more or less reliable and will tend to greater reliability.

However, if human cognition is utterly epistemically depraved, or the minimum probability required to bootstrap individual cognition to communal reliability too high, or the differential adaptive advantage accruing to true versus false beliefs too insubstantial, or all of the above, Plantinga's challenge will remain. If the naturalist were to aim to raise unaided and individually unreliable human cognition from its deep epistemic misery to respectability, she would indeed be begging the question. Although this wouldn't completely undermine her project – as every epistemology is ultimately circular (Russell, 1912: 112) – it should deflate her confidence that others would assume her epistemic stance.

2.3 Is Plantinga's distinction between 'guided' and 'unguided' evolution misguided?

Plantinga makes the claim that one can distinguish between what he calls 'guided' and 'unguided' evolution (Plantinga, 1994; Plantinga, 2011b). Guided evolution would be where evolutionary processes are 'guided' or 'orchestrated' by God, while on unguided evolution this wouldn't be the case. He claims that the idea of evolution being 'guided' or 'unguided' is not part of the relevant *science* – it's not part of the biology *as* biology – but are rather philosophical 'add-ons'. And the idea that God has guided the evolutionary process is consistent with evolutionary science. As Plantinga puts it:

> The scientific theory of evolution just as such is entirely compatible with the thought that God has guided and orchestrated the course of evolution, planned and directed it, in such a way as to achieve the ends he intends. Perhaps he causes the right mutations to arise at the right time; perhaps he preserves certain populations from extinction; perhaps he is active in many other ways. On the one hand… we have the scientific theory, and on the other…the claim that the course of evolution is not directed or guided or orchestrated by anyone… This claim, however… is no part of the scientific theory as such; it is instead a metaphysical or theological add-on (Plantinga, 2011b: 309).

Plantinga's claim that naturalism and theism are metaphysical 'add-ons' to the scientific theory of evolution depends on the nature or type of 'randomness' implied by the claim that *random* mutations are the main generative force driving the evolutionary process. If the 'randomness' involved in this developmental process is one of pure chance, it *would* preclude human origins from being the result of design. But, according to Plantinga, the randomness in biology – the random genetic mutations:

> [A]re random in the sense that they don't arise out of the organism's design plan and don't ordinarily play a role in its viability; perhaps they are also random in the sense that they are not predictable (Plantinga, 1994: 3).

Ernst Mayr, an eminent biologist, endorses this view when he notes that:

> When it is said that mutation or variation is random, the statement simply means that there is no correlation between the production of new genotypes and the adaptational needs of an organism in a given environment (Mayr, 1988: 99).

In other words, the science of evolution, specifically the science concerned with the process of genetic mutation, appears to imply – at most – that natural selection is not aimed at the direct benefit of any creature or population of creatures in which these mutations occur. Hence, the naturalist would be moving beyond the purely scientific evidence were she to make the claim that the evolutionary process is the result of pure chance, 'uncaused… or (not) arranged by God', *even were the process random in this way.* As Plantinga notes:

> ... [S]uppose the biologists, or others, *did* intend this stronger sense of 'random' (pure chance): then their theory (call it 'T') would indeed entail that human beings have not been designed by God. But T would not be more probable than not with respect to the evidence. For there would be an empirically equivalent theory (the theory that results from T by taking the weaker sense of 'random' and adding that God…orchestrated the mutations) that is inconsistent with T but as well supported by the evidence… (Plantinga, 1994: 3, his emphasis).

The claim Plantinga is making here is that the *empirical* evidence cannot be used to determine whether the randomness involved in the evolutionary process is ultimately that of pure chance or of the God-guided kind. For he argues that there could be two empirically *equivalent* theories that nonetheless disagree on whether God is or was involved or not.

Critics have responded that evolutionary naturalism *is* 'the standard *scientific* view of unguided evolution through purely natural mechanisms'(Boudry & Vlerick, 2014: 66, my emphasis). In other words, evolutionary science just *is* evolutionary naturalism, or more correctly, a proper subset thereof. Boudry agrees that the random mutations driving the process of evolution are random insofar as they are not to the benefit of the organism or its adaptational needs:

> "Random" in this context (genetic mutations) does not mean pure chance, but rather without foresight, or not necessarily concordant with the organisms adaptational needs (Boudry, 2013: 1219).

But he claims that:

> The thesis that mutations are random – i.e. unguided – rather than being a metaphysical afterthought, has been amply demonstrated and is nowadays accepted as the null hypothesis by evolutionary biologists. Experiment after experiment has shown that there is no evidence of non-random mutations arising because the organism "needs" them. Further, if some intelligent agent is triggering mutations after all, it seems he/she/it is causing precisely the kind and rate of mutations that one would expect if the process

> *were* entirely undirected (Boudry, 2013: 1220, his emphasis).

Recall that Plantinga makes the claim that God could have 'guide(d) and orchestrated the course of evolution, planned and directed it, in such a way as to achieve the ends he intends... (For example), 'perhaps he causes the right mutations to arise at the right time…' (Plantinga, 2011: 309). But wouldn't this imply that the mutations are to the (direct) benefit of the creatures or populations in which they occur? And isn't this *contra* his admission that the sense of randomness implied by the science of evolutionary theory is such that that the mutations 'don't arise out of the organism's design plan and don't ordinarily play a role in its viability… (Plantinga, 1994: 3)?

Not necessarily, for I see at least one way in which the claim that God is directing the process of evolution to his ends and the claim that genetic mutations are not to the 'benefit' or 'adaptational' needs of the creatures in which they occur can be reconciled. For example, God, being omniscient (on traditional Christian theism anyway), would have knowledge of all logical truths, including those about the future. Hence, he could choose from all logically possible worlds such that the mutations that occur in the world he chooses to create are consistent with the scientific evidence. In other words, God could create a world where mutations don't occur to the benefit of creatures *and* be directing the relevant evolutionary process after all. And clearly, this is a logical possibility. Still, as Boudry notes:

> If the bar for rational belief is lowered to mere logical possibility, and the demand for positive evidence dropped, then no holds are barred. Evolution (or gravity, plate tectonics, lightning…) could as well be directed by space aliens, Zeus, or the flying spaghetti monster… (Boudry, 2013: 1220).

To my mind, Plantinga's distinction between guided and unguided evolution holds, not as a result only of its logical possibility, but because there are independent reasons (i.e. independent of biology) to think that God may be upholding or directing nature's affairs after all – arguments from the 'fine-tuning' of the universe for example. Moreover, *pace* Boudry, his claim that 'the thesis that mutations are random – i.e. unguided – … has been amply *demonstrated*… by experiment after experiment' is not true (Boudry, 2013: 1220). For no empirical experiment can *definitively* prove that God isn't involved in the evolutionary process, and hence no empirical evidence can 'demonstrate' that pure chance runs the evolutionary show. Simply put, Plantinga's claim stands – both theism and naturalism are each ultimately metaphysical add-ons to the science of evolution.

On inspection, the objections raised in Section 2 lack the requisite force to trouble Plantinga – the evolutionary argument against naturalism *is* capable of 'launching' as it stands (*pace* Van Fraassen). Moreover, nothing of consequence results from Plantinga's alleged 'simplistic' framing of premise 1 – the argument wouldn't have to be aborted on account of such a 'simplistic framing' (were it simplistic). Finally, one can reasonably agree with Plantinga that evolutionary naturalism is *not* 'the standard scientific view of unguided evolution through purely natural mechanisms', but, like theism, a metaphysical addition to the relevant science.

3. The probability of the reliability of human cognitive faculties on naturalism and evolution

Recall that the engine of Plantinga's argument in support of premise 1 runs on the claim that evolution selects for adaptive behaviour, not belief (cf. Chapter 1). Further, he argues that on none of the three possibilities open to the naturalist in accounting for the link between belief and behaviour would the result be an epistemically happy one. In other words, were either of these three mutually exclusive and jointly exhaustive scenarios – (semantic epiphenomenalism, reductive physicalism, and non-reductive physicalism) – true, one shouldn't expect human cognition to be reliable (cf. Chapter 1).

More specifically, Plantinga claims that there's *no* reason on naturalism, evolution and either of semantic epiphenomenalism or non-reductive physicalism to think that human cognition would be reliable (Plantinga, 2011b: 330). For, on both, the link between belief and behaviour isn't causal, from which he concludes that evolution cannot select for belief content on either, and thus *a fortiori*, cannot select for reliable content. And, according to Plantinga, even if the link *were* causal, as on naturalism, evolution, and reductive physicalism, there would still be no reason to think that the beliefs so influenced will be in the direction of greater reliability (Plantinga, 2011a: 442).

Naturalists disagree. The essence of their various responses is that natural selection – *pace* Plantinga – *is* capable of selecting for belief whether the link between belief content and behaviour is causal or not (Boudry & Vlerick, 2014; Law, 2012; Ramsey, 2002). The arguments in support of these claims follow.

3.1 The probability of the reliability of human cognitive faculties on naturalism, evolution, and reductive physicalism

On naturalism, evolution, and reductive physicalism, Plantinga acknowledges that the 'link' between belief and behaviour is causal. Hence, natural selection can select for beliefs as a result of its ability to sample or sift neurophysiological structures that prove conducive to adaptive behaviour. But would (mostly) true beliefs be associated with neurophysiological properties that prove adaptive? Plantinga is not convinced:

> The neurology causes adaptive behaviour and also causes or determines belief content: but there's no reason to suppose that the content thus determined is true (Plantinga, 2011b: 327).

And:

> Take a neurophysiological property *P* that is in fact adaptive, and is also identical with the property of having content *p*: so far as adaptivity goes, it doesn't in general matter whether *p* is true or false. *P* would have the same effects, one thinks, if *p* were false. How *P* contributes to motor output doesn't depend (except in special cases) upon the truth-value of *p* (Plantinga, 2002: 218).

Naturalists agree that a given neurophysiological property and its associated belief – call it *P*-and-*p* – would lead

to the same behaviour whether *p* is true *or* false (Boudry & Vlerick, 2014: 71, 72).[30] In other words, given a *specific P*-and-*p*, *whatever p's* truth-value, there'll be some *specific* effect (behaviour) due to the causal link between *p* and the neurology with which it is identified (Boudry & Vlerick, 2014: 71, 72). However, the same behaviour will prove fitness enhancing or not depending on the *environment* within which this behaviour occurs. And it is here where the truth-value of *p will* come into its own; it will matter for the fitness of a creatures whether the *p* in the *P*-and-*p* combination is true or not (Ramsey, 2002: 17, 18; Boudry & Vlerick, 2014: 71, 72).

For example, suppose that the state of some investment banker's brain *P* is such that he believes that going camping will be a great way of reconnecting with nature amidst his feverish deal-making activities. Suppose he also believes that the river that runs through it – that part of nature with which he wants to reconnect – is perfectly safe for swimming (no rapids, no hippos, no crocodiles, and no other nasties of a predatorial or parasitic nature). Whether this particular investment banker "reconnects" with nature in a pleasurable way depends not *only* on the state of his brain or the content of his belief, but *also* on whether the river in question is as he takes it to be… safe. Hence, *pace* Plantinga, 'as (far) as adaptivity goes, it *will* in general matter whether *p* is true or false' (Boudry & Vlerick, 2014: 71, 72, their emphasis).

To my mind, naturalists are successful in showing that evolution, as Boudry and Vlerick puts it, 'does care about truth'. For creatures that form, hold to, and act on true beliefs will likely come to dominate the gene pool, as 'true beliefs (would on average be) better guides to behaviour than false ones' (Boudry & Vlerick, 2014: 68). Creatures whose beliefs accurately correspond to the relevant features of the environment within which they have to "behave" are likely to be more reproductively successful than those whose beliefs don't track those features as reliably.

3.1.1 Plantinga's 'belief-cum-desire' argument[31]

As noted, Plantinga agrees that evolution can select for both belief and behaviour on reductive physicalism (Plantinga, 2011b: 327). But, according to him, this doesn't give one reason to expect that the resultant beliefs would be true. For he argues that behaviour is not only the result of belief, but also desire (Plantinga, 1993: 225; Plantinga, 2011b: 327). Further, or so he claims, there are innumerable belief-desire combinations such that the resultant behaviour would prove adaptive, and further, as adaptive, had the fitness enhancing behaviour resulted from true beliefs (cf. Chapter 1; Plantinga, 1993: 225). Paul the hominid was introduced to show how this might happen (cf. Chapter 1).

Evan Fales argues that Plantinga's belief-desire scenario may work for simple perceptual belief-desire pairs or the odd inductive generalization, but it's not plausible in more realistic (complex) contexts (Fales, 2002: 51). He asks the reader to consider a deductive system of inference that employs true premises (true beliefs) versus one that employs false ones:

[30] The relevant 'association' being one of identity.

[31] The name "belief-cum-desire" is borrowed from Law (2011).

> *[T]rue* premises guarantee true conclusions"… a system that relies consistently upon true inputs to guide inference and action can employ general rules and hope to get things (i.e. action) right. But when a deductive argument employs false premises, the truth-value of the conclusion is *random* (Fales, 2002: 51, his emphasis).

Fales' point here is simply that sound conclusions inevitably follow when true premises are employed in logically-proper deductive arguments. On the other hand, one has no such guarantee when *false* premises serve as the inputs to such inferential schemas. In fact, one shouldn't be confident that such reasoning would generally lead to truth at all. As a result, Fales claims that there are no effective algorithms ('general rules') to implement Plantinga's schema – i.e. no evolved mechanism or *system* that would be capable of generating fitness-enhancing behaviour in an evolutionary realistic environment (Fales, 2002). And even were such a system or systems possible, there's reason to believe it would have to be more complex than a system that takes true inputs, appropriate desires, and as a result, outputs behaviour that would likely prove generally adaptive (Fales, 1996: 443).

Law tries to meet Fales' challenge. He considers whether evolution might – despite Fales' claims – have happened upon a cognitive *system* that could generate such false-but-adaptive belief-desire pairs (Law, 2011: 247 - 249). He entertains two possibilities; one, a system that generates truth-values *unreliably,* and two, one that does so *reliably*, but consistently exchanges truth for falsity. He then asks whether there are a set of desires such that the resultant belief-desire combinations can be expected to lead to (generally) adaptive behaviour (Law, 2011: 247, 248)? In what follows, I discuss and evaluate his attempts at answering this question.

3.1.1.2 Is there a desire or set of desires that could evolve given an *unreliable* belief generating mechanism?

Following Plantinga, one can say that an unreliable belief generating mechanism would be one that wouldn't result in its employer having mostly true beliefs – cf. Chapter 1. A system generating beliefs at a rate of one true belief for every false one would then by definition be unreliable. Law's question is whether a cognitive system of this nature can be fitted with some desire or set of desires such that the result would be creatures that behave in a generally adaptive manner (Law, 2011: 247, 248). To this end, he constructs a simple evolutionary scenario, fitting the relevant cognitive agents with an inferential mechanism that employs the fallacy of affirming-the-consequent. He assumes that such a cognitive system would be unreliable (Law, 2011: 247).[32]

Law asks us to consider the following sort of scenario (Law, 2011: 247). Suppose there is a married couple – Paul and Eileen – that reason in this peculiar way – affirming-the-consequent, as, and when, the need arises. For example, imagine the following scenario. Paul reasons as follows:

1. If hugging snakes is not safe, then hugging crocodiles is not safe.
2. Hugging crocodiles is not safe.

[32] The logical form of the fallacy of affirming-the-consequent is: if p then q, q, therefore p.

3. Therefore hugging snakes is not safe.

Eileen:

1. If hugging snakes is safe, then hugging crocodiles is safe.
2. Hugging crocodiles is safe.
3. Therefore hugging snakes is safe.

The question is: 'Is there a desire or set of desires that evolution could fit to an unreliable belief generating system such that both Paul and his wife could survive in this, and relevantly similar, contexts (Law, 2011: 248)?' In other words, can evolution "come up" with a desire that could be coupled with a system that generates mostly false beliefs such that the result is behaviour that proves generally adaptive? Let's see.

Consider Paul and Eileen's snake-and-crocodile situation once more. Suppose that both Paul and Eileen have the desire to die. Given his desire, and the relevant context, Paul will die. But Eileen won't. On the other hand, assuming that they don't want to die, Paul will remain among the living but his wife wouldn't. Hence, in this scenario – and in others relevantly like it – there appears to be no fixed desire or set of desires available such that they will *both* survive (Law, 2011: 247, 248). Still, when they desired to die, Eileen *did* survive, which means that there is at least *one* desire that can be fitted to an affirming-the-consequent rule such that the result isn't evolutionary deletion (at least in circumstances relevantly similar to the above).[33] But, as a demonstration that it is generally the case that a desire or set of desires can be fitted to an unreliable belief generating system such that the result is an evolutionary happy one, it fails. How so?

Firstly, it establishes only that an unreliable cognitive system can apply its algorithm (rule) to evolutionary success in a *singular* instance. It's a one-time application of a fallacious reasoning schema – employing false beliefs – that happened to result in an evolutionary amicable outcome. To be an evolutionary *possibility,* the mechanism needs to lead to evolutionary success *over time* – it needs to work on multiple occasions and in contexts where nature's demands are ever-changing. To be evolutionary *plausible,* it needs to be as evolutionary effective and efficient as a mechanism that takes true inputs, appropriate desires, and outputs fitness enhancing behaviour (more often than not).

As I will argue, an affirming-the-consequent-inferential system is only likely to work in very simple environmental contexts – such as Paul-and-Eileen's snake scenario. One reason is that such a system implements an invalid form of deductive reasoning. And in any form of such reasoning, the truth of the premises do *not* guarantee the truth of the conclusion. Indeed, even were *every* premise in *every* argument someone reasoning in this manner true, the truth of conclusion could *still* be false.[34] On the other hand, in any form of valid deductive

[33] Or Paul would have been the one to survive, if their desire was to live – i.e. *either* Paul *or* Eileen can survive given this context and the desire to live or die, but not *both.*

[34] For example: in an 'if... then' premise, the premise could be true whether the antecedent – e.g. 'If hugging snakes is not safe' – is true *or* false. Hence, premise 1 can be true, but the conclusion false. For example, if Paul concludes that hugging snakes is not safe, when it is.

reasoning, the truth of the premises guarantees the truth of the conclusion. It is thus highly likely – *ceteris paribus* – that a cognitive system implementing an invalid form of reasoning – e.g. affirming the consequent – will be inferior to one employing a valid one – e.g. modens tollens.

Secondly, it was assumed that *most* of the beliefs Paul and Eileen hold to are false, whether they are generated by a rule of inference or *not*. This means that most of the Paul and Eileen's basic beliefs – those not depending on others or formed on the basis of inference – would be false as well. Thus, not only would their cognitive system be implementing an invalid rule of inference, it would be doing so on defective raw material. I find it hard to believe that this kind of cognitive system would be an evolutionary possibility, especially in the sort of environmental context(s) humans are said to have evolved.

In illustration and in support of this claim, return to Paul and Eileen. Suppose it turned out that Paul survived the snake-and-crocodile-hugging scenario. Assume that he confronts a novel evolutionary challenge, one where he reasons as follows:

1. If I don't drink water, I will live.

2. I will live.

3. Therefore I won't drink water.

Given this reasoning, and assuming he desires to die, he will. Paul may have survived the crocodiles and the snakes, but he won't escape death by dehydration. However, given Paul's death wish and his tendency to believe mostly falsehoods, he *could* survive numerous rounds of evolutionary living. As far as I can see, this would only happen if two conditions are met: One, the content of Paul's beliefs should *reliably* be about states of affairs, which if acted upon, *won't* result in his death, or make it very unlikely. However, he must believe that they *would.* And, two, he should invariably have the desire to die. Given these two conditions, Paul will live. For he would reliably – but falsely – believe that acting on beliefs with this content will lead to his demise, when it (likely) wouldn't. However, the same could be said for a reliable cognitive system. It would employ mostly true beliefs, and when coupled with a more life-affirming desire or set of desires, would just as likely lead to survival in a new round of evolutionary living. The question is; 'Which would be the more likely to evolve, and why?'.

To my mind, a reliable cognitive system would be a far more likely evolutionary candidate. For it would arguably be far better at dealing with *novel* evolutionary challenges than its opponent. It would track and represent the salient features of novel evolutionary contexts as truthfully or accurately as its capacities for representation allows. On the other hand, an unreliable cognitive system would not do so – by definition. However, for Paul the "unreliable" to survive in novel contexts, his cognitive system would arguably need to be reliable after all. For it would reliably have to furnish him with beliefs whose content is mostly *false with respect to what's required for dying*, but mostly *true with respect to what's required for living*. In other words, Paul desires to die. But the only way he's not going to is if his cognitive faculties reliably misleads him with respect to what's required for dying. And reliably steering Paul away from death is pretty much the same as reliably steering him towards what may prove conducive to reproductive success. In short, Paul won't survive for long if his unreliable cognitive faculties

don't at least reliably represent those features of his environment which, if acted on, would likely result in his death. Evolution might opt for *pseudo* unreliable cognitive faculties, but it would be hard-pressed to select for cognitive mechanisms that are *truly* unreliable.

The lesson to be drawn from the arguments examined above is that there might be the odd circumstance in which someone reasoning by means of the fallacy of affirming-the-consequent would survive. But it would be truly remarkable, a stunning piece of cosmic serendipity, if he were to survive for any length of time in an evolutionary plausible context. So yes, it's possible, but vanishingly unlikely, that there is an evolutionary apt desire or set of desires that could be fitted to an unreliable reasoning mechanism that would result in reproductive success for the person that reasons in this fallacious manner. For anyone facing the challenge of survival in a complex environment, and who has to do so by forming beliefs and acting on desires, needs to be alive long enough to have a good chance of imprinting their progeny onto the future. However, reproduction, even for the most talented, takes some time and doing. As a consequence, there's little chance that such a system, were it to work, would do so as effectively and efficiently as one that reliably takes true inputs, appropriate desires, and as a result, outputs behaviour that would likely prove generally adaptive.

3.1.1.3 Is there a desire or set of desires that could be fitted by evolution to a belief-generating mechanism that *reliably* outputs falsities?

As shown, there doesn't appear to be a desire or set of desires that could be fitted to an unreliable belief generating system which would prove as effective and efficient as one that is generally reliable. But, asks Law, isn't it possible that a desire or set of desires can be fitted to a creature that reasons reliably, but consistently exchanges truth for falsity (Law, 2011: 248, 249)? More specifically, is there no desire or set of desires that could be fitted to a creature that reasons by a process of counter-induction – i.e. systematically exchanging truth for falsity – that would make its reproductive success likely (Law, 2011: 248, 249)?[35] According to him, there are good reasons to think not.

Law argues that (counter-inductive) inference coupled with a given desire may prove adaptive on *occasion*, but wouldn't work if a *series* of such counter-inductive inferences need to be made. For in any plausible evolutionary context, the ever-changing social and non-social dynamics of any recognizable human living would very likely often require a series of inferences. If this is true, then no person whose only inferential tool is counter-induction will last long (Law, 2011: 248, 249). In illustration, consider the following example.

Assume that some person has a desire to die. Next, suppose she reasons in the following manner; this is a poisonous plant and if I eat it I will live. That's a poisonous plant and if I eat it I will live. Hence, all plants are poisonous and if I eat any I will live. But I want to die. Hence I *won't* eat any poisonous plants. She will survive the aforementioned round of desiring and acting as she did (Law, 2011: 248, 249). But rewind the tape of her inferential life one frame. Now, suppose that the she reasons as follows; this is a non-poisonous plant and if I eat

[35] Whereas one would conclude (by induction) that if *every* cheese spoken to thus far didn't chalk, then the next cheese spoken to isn't going to chalk either, by counter induction, one would conclude from the same premise that the next cheese spoken to *will* chalk.

it, I will live. That's a non-poisonous plant and if I eat it, I will live. Eating non-poisonous plants will let me live. I want to die. Hence, I will *not* eat non-poisonous plants. She will avoid poisoning, but not starvation (assuming she's a strict vegetarian) (Law, 2011: 248, 249). Counter-inductive reasoning – as a procedure of inference leading to sufficiently and consistently adaptive behaviour – thus appears, if not impossible, unlikely to feature in any plausible account of the evolutionary origins and operation of human cognitive faculties (Law, 2011: 248, 249).

3.1.1.4 Laws conclusion: Fales' challenge cannot be met.

Recall that Law's aim was to try and meet Fales' challenge. And Fales' challenge was ultimately directed at Plantinga, where he claimed – *contra* Plantinga – that no cognitive *mechanisms* or *systems* could evolve that would be capable of generating fitness enhancing behaviour in an evolutionary realistic environment (Fales, 2002: 51). In trying to meet Fales' challenge, Law asked whether a desire or set of desires exists which could be fitted to either an unreliable affirming-the-consequent cognitive system or a system that systematically exchanges truth for falsity. On both counts, his answer was no. There is no desire or set of desires that can plausibly be fitted to a cognitive system either implementing the fallacy of affirming-the-consequent or a rule of counter-induction in which the result would be evolutionary plausible.

Note, however, that Law only considered one type of unreliable cognitive mechanism – one employing affirming-the-consequent. He hasn't shown that there aren't *any* unreliable cognitive systems that couldn't work. As far as I can see, there's no way of ruling out such a possibility, but to my mind, it would be vanishingly unlikely. For not only would such a system have to possible on *evolution*, but much more importantly, it would need to work well enough to be considered a plausible alternative to a reliable system. And I don't think there is any unreliable cognitive system that could meet the latter criterium.

Finally, what about the reliable counter-induction system? It's not unreliable, so perhaps it doesn't fall foul of either the possibility criterion or the plausibility criterion – neither been an evolutionary impossibility nor an implausibility? Still, I think Law's arguments against a system employing counter-induction is persuasive, but only because he assumed there was no *fixed* desire or set of desires that be fitted to such a mechanism. But couldn't a system implementing counter-induction be paired with a desire generating *mechanism* such that the result is a serious evolutionary contender? I don't think so.

For the problem with any type of procedural reasoning system that systematically generates falsities – e.g. one employing counter-induction – is that there are innumerably more ways of being wrong about the world than right. For example, Paul can rightly believe that there's a bird in the bush, assuming there is, or falsely believe that there isn't – i.e. truly believing p or falsely believing not-p (where p is the belief that there is a bird in the bush). Crucially, however, whereas *only* 'there is a bird in the bush' will make p true, there are a potentially *infinite* number of propositions that would make p false, or, equivalently, not-p true – i.e. Paul can believe there's a lion in the bush, *or* a donkey, *or* a playmate, *or*...

Hence, as far as I can see, any system that systematically outputs falsities is likely to lose its grip on features important to creatures' survival rather quickly, and thus be a cognitive mechanism unlikely to evolve. However,

were it to evolve, or at least be a plausible candidate for evolving, the *desire*-generating part of the system would need to be quite special. Firstly, it would need an extremely large repertoire of desires to augment the possibly countless ways in which the beliefs generated by the system could be wrong. Secondly, it would have to (dynamically) match the right desire or desires with the relevant belief or beliefs such that the resultant behaviour proves sufficiently adaptive. And if it would have to do all *that*, *it* would be doing all – or most of – the work one would normally think reserved for *belief*-generating modules, systems, or mechanisms. Hence, such a belief generating system would at best be extra and unnecessary baggage, or at worst, result in evolutionary deletion sooner rather than later. Further, such a system would likely be less efficient, and very likely less effective, than a system less encumbered.

3.1.1.5 Not by reasoning alone

Law is aware that reasoning alone will not give one confidence in thinking that the probability of the reliability of human cognition on naturalism, evolution, and reductive physicalism would be sufficiently high. The reason being that procedural reasoning needs to work in *conjunction* with other cognitive systems – such as memory and perception – which themselves need to be working sufficiently reliably to lead to the reasonable expectation that human cognitive faculties would be reliable (Law, 2011: 250). The argument that systemically unreliable memory or perceptual mechanisms are unlikely to evolve is largely the same as those with regards to procedural reasoning. On the odd occasion such unreliable faculties (of memory or perception) coupled with appropriate desires *could* work, but not *systematically,* and not in any realistic evolutionary context (Law, 2011: 250 - 254). What's more, Law argues that even if 'creatures possessed unreliable perceptual faculties… the probability of the reliability of human cognition might still be high' (Law, 2011: 254). He explains:

> [T]here arises the possibility – perhaps the probability – that the members of this species will be able to figure out that they are, to some extent, being systematically misled by those faculties (of perception). In which case, they may well adjust their beliefs accordingly. Their beliefs *would now reliably reflect reality, despite the fact that they possessed unreliable perceptual faculties.* If R is the reliability of their cognitive faculties *acting in tandem,* the probability of R might still be high, even if it was more probable than not that they possessed unreliable perceptual faculties… (Law, 2011: 254, his emphasis)

Fales and Law appear to have relegated Paul the hominid to an untimely death, an evolutionary miscarriage, or unicorn status. Plantinga's tiger-petting Paul *could* survive the odd round of evolutionary living, but not the continual probing or interrogation of a realistic evolutionary environment.

3.1.1.6 Unreliable but adaptive belief generating *mechanisms*

In response to Fales and Law's challenge to provide an example of a workable 'general algorithm' or an unreliable but adaptive cognitive *mechanism*, Plantinga resurrects and refashions Paul and his conspecifics. Suppose, suggests Plantinga, that Paul-and-clan 2.0 are creatures who believe that:

> [E]verything has been created by God… that everything is a *creature*, something created by God (and suppose that this is false)… Suppose further that their only way of referring to the various things in their environment is by (such) definite descriptions as the 'the tree creature before me' or 'the tiger creature approaching me'… [and take it] that all their beliefs are properly expressed by singular sentences whose subjects are definite descriptions expressing properties that entail the property of creaturehood… Suppose, finally, that their definite descriptions work the way Bertrand Russell thought definite descriptions work: 'The tallest man in Boston is wise', for example, abbreviates to 'There is exactly one tallest man in Boston, and it is wise.' Then from a naturalist perspective, all their beliefs are false. Yet these can still be adaptive: all they have to do is ascribe the right properties to the right 'creatures' (Plantinga, 2002: 260, his emphasis).

In other words, Plantinga's recipe for constructing 'deeply flawed' but adaptive cognitive faculties is to have the relevant creatures be *metaphysically*-mistaken Russellian semanticists. Why *metaphysically* mistaken? Because, what better way of ensuring that Paul-and-clan behave as adaptively as reliable "believers" than to maintain the accuracy of these creatures' mapping of the structural and temporal relations that obtain between the objects within their environment, while concomitantly slipping their beliefs about the substances so structured a poison pill. A belief can be wrong about the *substantial* or *metaphysical* nature of something, while accurately tracking the *structural* and *temporal* features or relations that hold between such substances.[36] It's the accurate tracking of the structural and temporal relations that hold between objects that allow creatures to survive, not the reliability of their beliefs concerning the metaphysical nature of said objects.

But, even if Paul and his troop *were* mapping the structural and temporal features of their environments accurately – so that the "part" of their beliefs about structural or temporal relations is accurate – their beliefs would *still* be false on Russell's logic of definite descriptions, as *that* to which they are ascribing structural or temporal features would be a *thing* or *object*, the nature of which they are getting wrong (by design). Simply put, even if Paul-and-clan are right about many or most of the structural or temporal relations that hold between objects in their environment, they are, by design, wrong about the *objects*, the metaphysical nature of which they are referring to falsely.

For example, suppose that Paul forms a belief about the location of a rock. Further, suppose he believes that the rock is alive – it's a creature of some sort. If Paul believes this, he may be right in thinking that the relevant rock is over there, but his belief concerning its location would nonetheless be false as no rock-creatures exist. He believes that there's a rock over there *and* that its alive, which is false.

In summary; if Russell's theory of definite descriptions is correct, Plantinga's response to Fales' challenge would be met.[37] There *would* be a system that takes false belief-desire pairs as inputs and outputs adaptive behaviour in

[36] At least with respect to 'medium-sized dry goods'.

[37] Of course, it could be the case that Russell's theory is incorrect. In fact, it's possible that our beliefs can be phrased in some externalist manner that allows for beliefs to be true even if they are driven by grossly false presuppositions.

a way comparable to more traditional or common-sensical cognitive mechanisms. For there appears no reason to think it would be less effective and efficient in its evolutionary operation than a system employing true beliefs and appropriate desires. It would be using the same, or very similar, cognitive machinery in the same way – generating beliefs with both a metaphysical and structural component combined with evolutionary appropriate desires. The only difference, and it seems one that makes no difference (relevant to cognitive efficiency and effectiveness), would be that the metaphysical beliefs generated by one cognitive system are false while those generated by the other true.

Having considered the arguments presented for and against the conclusion that the probability of the reliability of human cognition on naturalism, evolution, and reductive physicalism is low, what should one conclude? Should one expect human cognition to be reliable? Unreliable? Indeterminate?

Yes, Plantinga's metaphysically mistaken Russellians are an evolutionary possibility, and if their existence were likely on evolutionary naturalism, the probability of the reliability of human cognition would be low – very low. But the naturalist can acknowledge the possibility of the existence of such creatures without thinking such a scenario plausible or probable.

The fact that evolution can select for neurophysiological properties *P* on reductive physicalism and that a proprietary proposition is identified with the neurology so sifted provides one with good reason to think that the proposition is true. Natural selection can select for beliefs by selecting for behaviour, and the best explanation for differential adaptive behaviour is that the beliefs selected for are true – i.e. they are both the cause *of*, and most persuasive reason *why,* the resultant behaviour proves adaptive. Hence, one may conclude that the probability that human cognitive faculties are reliable on naturalism, evolution, and reductive physicalism is high. Or at least tends to reliability as individual human cognisors (slowly) learn that the reliability of the collective is more reliable, and useful for survival, than the thinking of any individual (see page 42 and 43 for how this might work).

3.2 The probability of the reliability of human cognitive faculties on naturalism, evolution, and either semantic epiphenomenalism or non-reductive physicalism.

3.2.1 The probability of the reliability of human cognition on semantic epiphenomenalism.

The arguments naturalists make in response to Plantinga's claim that human cognition is probably unreliable on semantic epiphenomenalism and non-reductive physicalism are very similar; hence the discussion under one heading. In effect, the critique naturalists offer with respect to Plantinga's argument against non-reductive physicalism can be seen as a subset of their critique of Plantinga's argument for semantic epiphenomenalism. The discussion to follow should therefore be seen in that light – a forceful critique of Plantinga's claim that naturalism, evolution, and semantic epiphenomenalism probably renders human cognition unreliable would by implication be a significant objection to his arguments against non-reductive physicalism.

Plantinga's charge that natural selection 'does not get to influence or modify the function (link) from

neurophysiological properties to content properties' as its 'just a matter of logic or causal law' appears most forceful with respect to semantic epiphenomenalism and non-reductive physicalism. For on both, beliefs *don't* enter the (physical) causal chain leading to behaviour (Plantinga, 2011b: 330). Moreover, if beliefs aren't causally effective, it seems that whether they are true or false wouldn't matter either (Plantinga, 2011a: 437, 444, 445). Hence, according to Plantinga, there wouldn't be any good reason to think that human cognition is probably reliable on naturalism, evolution, and either semantic epiphenomenalism or non-reductive physicalism (Plantinga, 2011a: 437, 444, 445). In fact, it appears that – at least on semantic epiphenomenalism – one would have good reason to expect human cognition to be *un*reliable.

Recall that Plantinga thinks that the link or relation between belief and behaviour on semantic epiphenomenalism, or more accurately, the relation between belief and the neurology that causes behaviour, is maximally permissive with respect to content (cf. Chapter 1). For example, on semantic epiphenomenalism, *different* beliefs can be associated with the *same* neurophysiological properties; the same neurology can support the belief that the cat is on the mat *and* the belief that it isn't.[38] Given that neurological properties are the sole cause of the relevant person's behaviour, it follows that different beliefs can lead to the same behaviour. Plantinga makes the point as follows:

> [I]t is exceedingly difficult to see… how they (beliefs) can enter (causal chains) *by virtue of their content: a given belief it seems, would have had the same causal impact on behaviour if it had had the same (physical) properties, but different content* (Plantinga, 2011a: 436, his italics).

Further, if the content of belief makes no causal difference, and the same neurophysiology can support beliefs with different content, it follows that the content of any given belief in any behavioural context doesn't have to be about the relevant features in which that behaviour occurs. In short, on semantic epiphenomenalism, the content of belief has no care in, or about, the world. As Plantinga puts it:

> Those beliefs ("associated" with certain neurophysiological properties) need not be so much as *about* the objects involved in the states of affairs causing the subvening properties. They could be about *anything* (Plantinga & Tooley, 2008: 232, his italics).

Plantinga argues from analogy in support of these claims – that beliefs are not only causally impotent, but their content – and *a fortiori* their truth-values – irrelevant to a creature's behaviour. He asks us to imagine an opera

---

[38] In other words, on semantic epiphenomenalism the content of a belief is *not* fixed or determined by its physical realizers. But on non-reductive physicalism it is. Given the supervenience relation, a belief's content *will* be fixed or determined by its physical realizers (whenever and wherever they are instantiated). Plantinga:

> [I]f content *supervenes* on neurophysiological properties... then it won't be possible... (for) the same neurophysiological properties ... (to have) different content (Plantinga, 2002: 214).

Further, on this view – unlike non-reductive physicalism – there *doesn't* need to be a change in the physical realizers of a belief if the content of that belief changes.

singer who is able to (and does) break a crystal glass as a result of her singing.[39] He claims that she doesn't break the glass as a result of what she is singing *about,* but rather as a result of the *physical* properties of her voice; the *content* of her singing appears causally irrelevant (Plantinga, 2002: 214). The same moral applies to a brick breaking a window. Whatever its colour, its physical properties (momentum and such) does the deed, not its colour. A white brick will break a window as effectively as the same brick by any other colour (Plantinga, 2002: 218).

In essence, Plantinga's point is this: if beliefs don't enter the causal chain leading to behaviour as a result of their *content*, then evolution cannot select for that or any other content. And if evolution can't do that, then *a fortiori*, it wouldn't be able to select for *true* content either. Simply put, if evolution can't affect belief content directly, and if there's no logical, metaphysical or nomological rule that ties the content of belief to its neurophysiological properties in an appropriate way, there wouldn't be any reason to expect cognitive faculties functioning in this manner to be reliable. He concludes:

> If semantic epiphenomenalism were true, it would be an enormous cosmic coincidence, a stunning piece of not-to-be-expected serendipity, if modification of behaviour in the direction of fitness also modified belief-production in the direction of greater reliability (Plantinga, 2011a: 437).

In response, Law (2012) claims that even if there are no *causal* links between belief and behaviour (on semantic epiphenomenalism), there would likely be *conceptual* ones (Law, 2012: 41). And if these conceptual links tie specific belief content to a certain belief structure or set of such structures, evolution *would – pace* Plantinga – be able to select for different content as it selects for different neural structures (Law, 2012: 44, 45). In other words, even were semantic epiphenomenalism true, it's possible that different – but specific – belief content can be conceptually linked to different and differentially fitness-enhancing belief structures. And given that evolution can select for the latter (belief structures), it will be able select for the former (belief content) by conceptual association (Law, 2012: 41, 44).

But why think there are such conceptual constraints? Law, like Plantinga, appeals to intuition. He argues that assuming that such constraints exist would make sense of our natural inclination to think that the content of someone's belief is not entirely independent of that person's behaviour. We make sense of others behaviour by assuming that their behaviour, like ours, is informed by beliefs aimed at fulfilling certain desires. Law explains:

> We can know *a priori*, solely on the basis of conceptual reflection, that, *ceteris paribus*, the fact that a belief/neural structure causes that behaviour [a person walking five miles south] in that situation [the person is thirsty] significantly raises the probability that it has the content there's water five miles south. Among the various candidates for being the semantic content of the belief/neural structure in question, the content that there's water five miles south will rank fairly high on the list (Law, 2012: 45).

[39] Plantinga borrows the opera-singer (analogy) from Dretske (Dretske, 1988: 80).

And:

> It seems intuitively obvious to many of us that belief content is not entirely conceptually independent of behavioural output: *that one cannot plug any old belief content into any old neural structure (or soul-stuff structure, or whatever)* entirely independently of its behavioural output. That intuition would appear to be, philosophically speaking, largely pre-theoretical. It cannot easily be dismissed by Plantinga as a product of some prior theoretical bias towards naturalism and/or materialism (Law, 2012: 48, my emphasis).

Crucially, Law doesn't assume that reductive or non-reductive physicalism is true – that the content of belief can be reduced to, or supervenes on, the physical. The conceptual constraints he claims are likely to exist are substrate neutral; the same conceptual constraints would likely hold no matter what the substantial make-up of the relevant behaviour-causing structure or structures may be. As he puts it:

> [T]o suggest that such conceptual constraints on belief content exist is not, of course, to presuppose that beliefs are neural structures or that materialism is true. Let's suppose, for the sake of argument, that substance dualism is true and that beliefs are not neural structures, but soul-stuff structures. Then my suggestion is that we may be able to know on the basis of a little conceptual reflection that if beliefs are soul-stuff structures, and if a given soul-stuff structure in combination with a strong desire for water typically results in subjects walking five miles south, then, *ceteris paribus,* that soul-stuff structure is quite likely to have the content that there's water five miles south, and is rather unlikely to have the content that there's water five miles north (Law, 2012: 46).

As noted, the take-home message of the above discussion is this: different – but specific – belief content can be conceptionally associated with different, and differentially successful behaviour-causing structures. Hence, by selecting for differentially adaptive structures, evolution would be selecting for content, even if belief content – *qua* content – is causally sterile. And if one assumes further that true beliefs are on average better guides to behaviour than false ones, then evolution would likely 'mould belief in the direction of greater reliability'. Hence, even were semantic epiphenomenalism true (on naturalism and evolution), one shouldn't find the trustworthiness of human cognition surprising.

Law's argument carries some force – especially given that is doesn't rely on naturalistic assumptions regarding the metaphysics of human nature or the nature of meaning. Whatever humans ultimately are – whether souls, brains, or something else – it's intuitive to conclude that not just any belief can be righty ascribed to anyone in any context. Content restrictions apply. And as far as I can see, the best reason to think that such conceptual or content restrictions apply is their usefulness (and indispensability) in making sense of human social interaction.

3.2.2 The probability of the reliability of human cognition on non-reductive physicalism

If Law's argument holds with respect to semantic epiphenomenalism, then it will likely hold with respect to non-reductive physicalism. For, as Law has convincingly shown, the metaphysical nature of the link between belief and behaviour is unimportant. In determining whether one belief ascription is sensible vis-à-vis another it

wouldn't matter if the relevant beliefs are being ascribed to souls or brains or something else. Boudry and Vlerick makes the point as follows:

> As long as NP states are associated with specific belief contents, either causally *or conceptually*, belief content is within the reach of natural selection, and Plantinga's EAAN fails. The precise *relation* between belief content and NP properties is irrelevant to Plantinga's argument (Boudry & Vlerick, 2014: 69, 70, my emphasis).

But suppose Law's argument proves unsuccessful – that it doesn't establish that there are such conceptual constraints, even if there were. In other words, suppose that the intuition that motivates Law's conclusion that there are likely such constraints proves weaker than Plantinga's intuition that it is 'exceedingly difficult to see' how beliefs 'can enter (causal chains) *by virtue of their content'* (Plantinga, 2011a: 436, his emphasis). And that it 'seems' that '*a given belief would have had the same causal impact on behaviour if it had had the same (physical) properties, but different content'* (Plantinga, 2011a: 436, his emphasis). If this is true – if Law's argument fails – it would appear that Plantinga would be right in claiming that we should expect human cognition to be unreliable on naturalism, evolution, and semantic epiphenomenalism. For if *any* content will do as well as any other from an adaptive point of view, then any content – whether true or false – will also do. Hence, according to Plantinga, there wouldn't be any reason to expect human cognition to be reliable on naturalism, evolution, and semantic epiphenomenalism (Plantinga, 2011a: 437).

But what about the expected reliability of human cognition on naturalism, evolution, and non-reductive physicalism? Would one likewise have no (good) reason to expect human cognition to be reliable on the latter, as Plantinga claims (Plantinga, 2011a: 444, 445)? Boudry and Vlerick don't think so. In response, they make two claims; firstly, they point out that on non-reductive physicalism different beliefs *cannot* be associated with the same neurophysiology (by supervenience) (Boudry & Vlerick, 2014: 70). Hence, as evolution selects for different belief structures, it would be selecting for (specific) belief content, which is the first step in it possibly selecting for reliable content.

And secondly, they claim that not just any content can be associated with the same-behaviour causing neurophysiology – as semantic epiphenomenalism appears to allow – on pain of incoherence (Boudry & Vlerick, 2014: 71). In support of their charge that semantic epiphenomenalism is incoherent, they appeal to functionalism, which – in the philosophy of mind – is the idea that a belief is *defined* by its causal or functional role (within a broader causal context).[40] To highlight how functionalism would render semantic epiphenomenalism incoherent,

[40] Note that functionalism is consistent with non-reductive physicalism in that the same causal role or function can be played by or fulfilled by different physical realizers. Both an Apple ('Mac') and a personal computer ('PC') – i.e. different physical systems – can run the same software – e.g. Windows – and can thus perform the same functions – e.g. those that Windows can. On non-reductive physicalism, the idea that the same belief can be realized by different physical systems – e.g. human brains or alien brains – is known as the concept of multiple realizability (Bickle, 2019).

they construct a counter-analogy of their own (in response to Plantinga's 'opera-singer' and 'brick' analogies, see page 55 and 56).

3.2.2.1 Non-reductive physicalism and functionalism

As noted, Boudry and Vlerick make the point that different belief content *cannot* be associated with the same neurophysiology (by supervenience) (Boudry & Vlerick, 2014: 70). This means that as evolution selects for specific belief structures – the subvening neurophysical 'realizers' of belief states – it would be selecting for specific belief content by supervening association. And this opens the possibility that evolution would select for adaptive belief content as it selects for the adaptive neurophysiology with which that content would inevitably covary. Further, if adaptive belief content is more likely to be adaptive *because* it is true, then evolution *would* be caring for the truth of creatures' beliefs by caring for their survival. The intuition that adaptive belief content is likely to be associated with adaptive neurophysiology – at least more often than not – is attractive. But perhaps it's not quite powerful enough to justifiably secure the thought that adaptive belief content is likely to be associated with adaptive neurophysiology. To add force to this intuition the non-reductive physicalist could appeal to functionalism. For, on functionalism, as we shall see, there is a good reason for thinking that evolution would likely select for adaptive belief content as it selects for fitness enhancing behaviour.

3.2.2.2 The non-reductive physicalist's friend: functionalism

As noted, on functionalism, a belief is *defined* by its causal or functional role (within a larger causal or functional context). In other words, the meaning of a belief is given by its causal or functional role. Hence, it wouldn't make sense to say that a different belief can be associated with the same "role-playing" neurophysiology. If a belief is defined by its causal or functional role, it cannot be *that* belief if it doesn't play *that* role in terms of which it is defined.[41] Boudry and Vlerick explain:

> Believing X means being prepared to take certain courses of action, when conjoined with a set of desires and other beliefs... If some NP property does not dispose us, *ceteris paribus*, to the behaviour that we would expect from some belief state X, *then it cannot be that belief state X.* For instance, if the belief 'this snake is dangerous' does not produce flight behaviour in the presence of a snake, when coupled with the desire 'I don't want to get bitten by a poisonous snake', and with other background beliefs, such as 'if I run away from X, then X is less likely to hurt me', then something must have gone wrong in our belief ascription. Belief contents are not arbitrary labels that one can tag to NP properties at will, as Plantinga supposes. The counterfactual claim that some NP property would have the same effect 'even if it had quite different content', therefore, is incoherent (Boudry & Vlerick, 2014: 71, my emphasis).

[41] Suppose belief *p* plays causal role R. Hence, by definition, the meaning of *p* is given by R. Further, whatever the content of belief *p* might be, it isn't *not-p*. If this is true, then *p* and *not-p* must play different causal roles, given that the meaning of *p* is defined by its causal role. If this is the case, it wouldn't make sense to suppose that *p* and *not-p* can play the *same* causal role. If one has identified what Kim calls the physical realizers that play the causal role specified – i.e. R – one cannot label or ascribe, on pain of incoherence, that *not-p* (or any other belief) is playing that role. *That* role – R (at that time and in that context) – is *p's* alone (Kim, 2005: 101).

In support of their claim that Plantinga is playing an unjustified, and in fact, incoherent game of "pin-the-donkey" with meaning, they construct an analogy of their own (in response to his opera-singer and window-breaking analogies).They acknowledge that balls don't break windows as a result of their colour (or their being a birthday present) – that there is no causal or conceptual link between a ball being coloured or it being a birthday present and its breaking a window. But they deny that the same holds with respect to belief content and its neurophysiology. Their counter-analogy runs as follows:

> [C]onsider an old-fashioned calculator. When I press some keys, numbers appear on the screen. Now suppose that Plantinga would come along and say that 'the calculator would produce the same result if the keys had different meaning'. By way of demonstration, he removes the labels from keys 3 and 4 and switches them: 'see, the calculator still produces the correct results'. But it should be obvious that the inscription or label on the key is not where meaning resides. The pressure applied to the key relays a signal on an electronic circuit, regardless of the inscriptions on the key. *The key represents number 3 in virtue of the fact that it is connected with the device in a certain manner*, such that it produces the expected results on the screen when pressed in combination with other keys. The manufacturer has just conveniently arranged my keypad in such a way that the inscriptions correspond to the internal representations, in this case numbers and mathematical concepts. If somebody secretly swapped the key labels, this will soon be discovered... (Boudry & Vlerick, 2014: 72, 73, my emphasis).

As you can see, their functionalism is apparent – a belief's contents, its meaning, is defined, and can be aptly explained, by its functional role (Kim, 2011: 169). It wouldn't make sense to ascribe a different content to a belief that stands in a certain well-defined causal role, for the causal role it plays *is* its meaning. Hence, given that one has successfully identified the causal role that a belief plays in a greater causal context, one has successfully pinned down its meaning. Hence, according to Boudry and Vlerick, to suggest that it can have any other meaning given that it plays *that* or *this* causal role would be incoherent (Boudry & Vlerick, 2014: 71; see note 39). And this fact, if it is a fact, would enable evolution to select for belief content as it selects among those causal roles that prove fitness enhancing. Moreover, if true beliefs are on average better guides to behaviour than false ones, the reliability of belief will be likewise be within the powers of evolution to select for. Hence, if it's justifiable that a belief's content can be defined by its causal role, the non-reductive physicalist would have an additional reason – in addition's to Law's intuitive conceptual constraints – to think that human cognition would likely be reliable on naturalism, evolution, and non-reductive physicalism.

Briefly, Boudry and Vlerick's aim has been to show that – on non-reductive physicalism – there are good reasons to think that the sort of conceptual constraints Law thinks exist, really do. They made two claims: the first is that on non-reductive physicalism, different belief contents cannot be associated with the same behaviour-causing neurophysiology. And hence, as evolution selects for different neurophysiologies, it would be selecting for different content. Secondly, they claimed that if a belief is defined by its causal role (within a broader causal or functional context), then it wouldn't make sense to think that just any content can be associated with any behaviour-causing neurophysiology. Thus, there is good reason, at least given functionalism, to think that the sorts of conceptual constraints Law think exists, do, and that evolution would be able to select for belief content as it selects for belief structure. If one adds the plausible assumption that true beliefs are on average better guides

to successful behaviour than false ones, then human cognition would likely be reliable on naturalism, evolution, and non-reductive physicalism.

If the arguments presented and discussed in Section 3 are successful, then premise 1 of the evolutionary argument against naturalism would be false. *Pace* Plantinga, human cognition would likely be reliable on naturalism and evolution.

But one might rightly ask why only the expected reliability of human cognition on each of the three proposed links between belief and behaviour were discussed, and not the probability that each scenario would be true (on naturalism and evolution)? For to determine the probability of the reliability of human cognition on naturalism and evolution it was shown – cf. Chapter 1– that one needs to consider *both* the probability of the reliability of human cognition on each of the relevant scenarios *and* the probability of those scenarios being true. In other words, the expected reliability of human cognition on naturalism and evolution is a function of *two* elements – the probability of the reliability of human cognition on each of the three proposed links between belief and behaviour *and* the probability of each scenario being true (on naturalism and evolution). So why only discuss the former when it appears that the latter also requires consideration?

Ordinarily, in any expected value calculation, one cannot ignore the probabilities attributed to either of the relevant scenarios happening, but in premise 1 of the evolutionary argument this isn't necessary. The reason is that on each possible scenario – semantic epiphenomenalism, non-reductive physicalism, and reductive physicalism – the probability that human cognition is reliable is high or sufficiently high. Hence, it doesn't matter what their probabilities of being true may be. On *each* scenario (and hence their sum), human cognition would likely be reliable. Simply put, *each* of the three terms expressed in the disaggregated probability equation would, in their own way, inevitably *sum* to a value that would establish that human cognitive faculties are likely reliable on naturalism and evolution.

4. Conclusion

In sections 1 and 2, the discussion focussed on a number of naturalist objections to Plantinga's framing of the probability thesis and his characterization of each of the concepts employed therein. The claims were that Plantinga's characterization of human cognitive reliability is overly simplistic, that his distinction between 'guided' and 'unguided' evolution is confused, and that his conception of naturalism is empty.

On closer inspection, it was argued that none of these charges – bar Plantinga's supposed 'simplistic' framing of human cognitive reliability – carry much force. Naturalists attempted to show, without begging the question, that even though the *individual's* cognitive reliability may likely be compromised, if it is, that it would still be reasonable to expect the *collective* to be cognitively reliable.

In Section 3, the question was whether (individual) human cognition would likely be reliable on naturalism and evolution and either of semantic epiphenomenalism, non-reductive physicalism, and reductive physicalism. On semantic epiphenomenalism, it appeared, initially, that the answer would be a straightforward no; belief content

plays no causal role in behaviour. Hence, there would seem to be little or no reason to think that human cognition would likely be trustworthy. However, Law convincingly showed that even though belief content plays no causal role on semantic epiphenomenalism, there would likely be conceptual constraints restricting what content can rightly be associated or ascribed to any behaviour-causing neurophysiological structure or structures.

On non-reductive physicalism, a proprietary belief is necessarily associated with any given neurophysiological property. Hence it makes sense – at least from a belief ascription perspective – to think that a *given* belief coupled with the appropriate desire leads a creature to behave *so-and-so* rather than *that-or-the-other* (in its environmental context at time *t*). Indeed, on functionalism, the content of a belief would be *defined* by its causal role. Hence it *cannot* be *that-or-the-other* belief if it doesn't play *that-or-the-other* causal role. On functionalism, it would be incoherent to suggest otherwise.

Moreover, the conceptual constraints that Law argues would hold with respect to semantic epiphenomenalism would apply to non-reductive physicalism as well. For the nature of the link between belief and behaviour is irrelevant to their possibly being conceptual constraints on belief content. Finally, given the plausible assumption that true beliefs are on average better guides to evolutionary successful behaviour than their opposites, it's reasonable to think that human cognition would likely tend to reliability on naturalism, evolution, and non-reductive physicalism.

The same line of argument was used to support the claim that human cognition would likely be reliable on naturalism, evolution, and reductive physicalism. However, on reductive physicalism, the argument carries more force. As belief content isn't merely *associated* with some physical (neurophysiological) property on reductive physical, but *identical* or *reducible* to it, natural selection would have greater purchasing power on selecting for belief content as it selects for fitness-enhancing neurophysiology.

In Chapter 3, the question will no longer be whether human cognition is probably reliable or unreliable in general, but which of naturalism or theism makes most sense of the evidence with respect to the distribution of human cognitive reliability among its different faculties, modules, and subject matters. In other words, *whatever* the absolute value or range of values of the probability of human cognition on naturalism and evolution (or theism and evolution), the question and discussion to follow will be concerned with how that reliability or unreliability is *distributed,* and what best explains it.

CHAPTER 3: EVOLUTIONARY NATURALISM ON THE DISTRIBUTION OF HUMAN COGNITIVE RELIABILITY

1. Introduction

Thus far, the arguments presented and discussed were aimed at discussing the possibility that – on naturalism and evolution – human cognition is probably either reliable (naturalists), or unreliable (Plantinga). In what follows, the *magnitude* of the probability of the reliability of human cognition, whatever that value may be, is less important than how that cognitive reliability – large or small – is *distributed*. In other words, *given* the magnitude of the probability of human cognitive reliability, what can one say with respect to the reliability of *specific* cognitive faculties or modules? For example, how reliable is human memory, human sensory perception, or human reasoning? Relatedly, concerning which subject matters should one find or expect their deliverances credible?

Notice that the question is no longer focused on the *probability* of the reliability of human cognition, but about its *de facto* reliability – i.e. cognitive reliability in *this* world. In other words, the question takes the world as given, and asks what the empirical evidence shows with respect to the reliability of human cognition, ignoring for the moment whether the world is one where evolutionary naturalism or theistic evolution is true. The empirical findings relevant to answering this question are presented in section's 2 and 3.

Given the evidence – on the distribution, shape, or contours of human cognitive reliability – the focus will shift to which of evolutionary naturalism or theistic evolution (theism and evolution) best explains these findings. Naturalists claim that the evidence finds a welcome and comfortable home within an unguided evolutionary framework, whereas the same cannot be said of theistic evolution. The arguments in support of this claim will be presented and evaluated in section 4, and in Chapter 4.

2. The empirical evidence with respect to human cognitive reliability: An overview

During the latter half of the twentieth century, cognitive scientists have shown that the working and products of human cognition are neither as transparent nor as infallible as had previously been supposed (Childers, 2011; Kahneman, 2011). As it turns out, the products of the human mind – its sensory perceptions, inferential practices, the testimony of memory, and introspective reliability – deviates from the strict norms of rationality – i.e. the strictures of logic and the 'norms' of probability (Boudry, Vlerick, & McKay, 2015: 85, 86; Kahneman & Tversky, 1973: 237). Moreover, the products of these faculties or capacities have been shown, in specific circumstances, to be inaccurate in representing empirical reality as measured by contemporary science (Kahneman, 2011; Childers, 2011). In other words, when the intersubjective standards and research protocols of the scientific establishment – peer review, repeatable experiments, (generally) agreed upon metrics of measurement, and so forth – are employed to measure or determine the reliability of individual human cognition, it has proved less reliable than one might have initially supposed.

The extent of the evidential basis for the claim that human cognition is less than optimally reliable precludes all

but a cursory overview.[42] However, the findings canvassed here should be sufficient to show that it's well supported.

2.1 Human cognitive reliability: the evidence

Human sensory perception can be fooled. Human reasoning (often) sacrifices accuracy for other goods like speed or ease of processing (Haselton, Nettle, & Murray, 2005; Kahneman, 2011). Human memory isn't the equivalent of a reliable audio-visual recording device, but takes liberties in its testimony (Loftus, 1974a; Loftus, 1974b; Childers, 2011). Further, evidence suggests that humans aren't particularly good at identifying the *real* reasons for their actions (Childers, 2011). Finally, individuals certainly appear to have serious difficulty in coming to reliable terms with subject matters far removed from those salient to matters of survival and reproductive success (Bourget & Chalmers, 2014).

Note that the reliability referred to, and discussed here, doesn't refer in any way to the *acuity* or *resolution* of the relevant human sensory apparatus. For example, the human sense of sight will be inaccurate when having to judge the (true) empirical features of the very small, the very faint, or the very far away. Human hearing won't detect the presence of noises that are too soft; neither will the sense of smell or that of taste detect chemical molecules whose concentrations are too low. Finally, the acuity of human tactile perception will prove too coarse-gained to detect the presence, or track the movement, of an object whose tactile impressions are too soft. In other words, the human senses have a proper range and domain of operation within which – if they're not malfunctioning – they ought to represent the empirical features of the relevant environment (more or less) reliably. Simply put, they won't reliably, or unreliably, represent what they cannot see, hear, smell, taste, or feel. The evidence to be discussed doesn't involve what is been *missed* by the senses, but rather what is *within* their range of proper functioning, but represented falsely.

2.1.1 Human sensory-perceptual reliability

Each of the five human sensory modalities – sight, hearing, feeling, smell, and taste – can be fooled *systematically*. More specifically, human perceptual systems that have evolved to be (sufficiently) reliable in their natural contexts have been found wanting when purposefully prodded to function in artificial environments (Haselton *et al*., 2005). This doesn't mean that humans are perfectly reliable "perceivers" in their native evolutionary contexts, only that their perceptual faculties' subtle deviations from reliably representing empirical reality can be made salient when pressed to perform in carefully constructed artificial environments.

For example, psychologists have fooled human visual perception by constructing contexts that take advantage of

[42] See Kahneman & Tversky (1973); Kahneman & Tversky (1974); Ariely (2010); and Kahneman (2011).

the human tendency to use perspectival cues to judge the size of objects (Kahneman, 2011). In this way, the visual system has been fooled into inferring that two figures are of different sizes when they're not, or that two lines are of the different lengths, when their objective measure is the same (Kahneman, 2011). Further, it has been shown that subjects tend to judge the height of cliffs differently depending on whether they're viewing these objects from the top or the bottom – e.g. subjects judge cliffs to be higher when viewed from the top than from the bottom (Haselton *et al*., 2005).

On a more entertaining note, magicians have been particularly adept at intentionally tricking human sight – exploiting its "blind spots" as it were. Coins and cards have been shown to disappear, the more accomplished appear able to walk on water, defy gravity, or render substantial material structures invisible, despite the audience's (presumably) rapt attention. For many, seeing may be required for believing, but alas, not (always) as reliably as they may suppose.

An example of hearing gone "wrong" is a phenomenon called auditory looming. This refers to the tendency of 'people to judge a sound that is rising in intensity to be closer and approaching more rapidly than an *equidistant* sound that is falling in intensity' (Haselton *et al*., 2005: 973). Further, research subjects have been shown to judge an 'approaching sound source to be closer by than a receding one, when in fact the sounds were located at distances equally far away' (Haselton *et al*., 2005: 973).

In certain contexts, smell doesn't reliably track matters olfactory within its range of proper functioning either. Stevenson has shown that human olfaction can suffer from two types of illusion. Firstly, there are instances 'where the same stimulus results in different percepts, and cases 'where different stimuli result in the same percept' (Stevenson, 2011: 1887). An example of the first – same stimulus, different percepts – has been observed to happen when subjects rated the pleasantness of a smell differently depending on whether they were initially cued or primed with a negative stimulus followed by a positive one, or vice versa (Stevenson, 2011: 1889, 1890).

For instance, when initially primed by the label 'toilet cleaner' and then asked to judge the pleasantness of a pine odour when presented with the label 'Christmas tree', subjects judged the shift or variance in the pleasantness of the pine odour to be *less* than if the order of the presentation of the stimuli were reversed – i.e. if the label marked 'Christmas tree' was followed by the label marked 'Toilet cleaner' (Stevenson, 2011: 1890). An example of the second type – different stimulus, same percept – has been shown to occur when subjects have reported the presence of the same smell – same percept – even though the chemical composition of the molecules present or available to the subjects' olfactory apparatus were different (Stevenson, 2011: 1892).

The human sense of taste is similarly untrustworthy in certain contexts; it too can misperceive. For example, it has been shown that the colour of a wine glass affects the wine's perceived quality (Ross, Bohlscheid, & Weller, 2008). A panel of consumers liked or enjoyed the taste of the same red wine more when drunk from a blue-tinged glass than a clear glass – given the same lighting conditions (Ross *et al*., 2008). Similarly, soft drinks served from a "colder-coloured" blue glass were perceived to be more thirst quenching than the same drink served in a "warmer-coloured" yellow glass (Spence, 2011: 101). Moreover, in experiments where café lattes were served in white mugs, they were judged to be less sweet – although more intense – than lattes served in a glass or blue

coloured mug (Van Doorn, Wuillemin, & Spence, 2014).

Finally, and to my mind, one of the most fascinating sensory-perceptual illusions is that of *induced* out-of-body experiences (OBE's) (Ehrsson, 2007). And no – these aren't (clearly) near-death experiences (NDE's) – at least insofar as the variability of the subjects' states of health before and after the experiments or experiences are concerned.[43] Further, this experiment also demonstrates that even human sensory modalities working in concert can be led astray – sight and touch in this case.

The basic experimental setup demonstrating the presence of this illusion involved a subject sitting on a chair wearing a head-mounted video display. The inputs to the display were two video cameras situated two meters behind the subject – simulating the perspective of someone viewing her from behind. The experimenter stood beside her – *in* her field of vision as captured by the cameras – i.e. from the visual perspective of the "someone" behind the subject, in this case that someone being her. The experimenter would then use two plastic rods, one to stroke her "real" chest (out of view of the cameras and hence out of her view), while simultaneously stroking her "virtual" chest in the cameras' field of vision, and hence within her (virtual) visual field (Ehrsson, 2007: 1048). This induced the feeling – as reported by the subject or subjects – that they were outside their bodies (Ehrsson, 2007: 1048). The putative explanation offered for this phenomena is that the brain has to makes sense of conflicting data – ocular inputs leading the subject to *see* herself being stroked in one location but *feel* herself being stroked in another (two meters in front of her). In trying to make sense of the conflicting data, the brain creates the illusion that the subject is outside her body (Blanke & Dieguez, 2009).

2.1.2 Human memory and mental transparency

Human memory, the human cognitive capacity to form, retain, and retrieve memories, isn't the equivalent of a reliable recording device (Loftus, 1974; Loftus, 1978; Loftus, 1997; Loftus, 2005; Childers, 2011). Human memory isn't nearly as trustworthy in capturing and retaining the audio-visuals of a creature's past as one might initially think. False memory and the effects of misinformation are two well-established examples of the license that the faculty of memory can take with respect to the past.

False memory is the phenomena whereby individuals have been shown or "induced" (to tragic effect on occasion) to remember events in their past – autobiographical events – that *didn't* happen (Loftus, 1997: 70). In an experiment carried out by Loftus (1997), subjects were presented with a list of events – most of which were events that did occur – but by design included events that didn't (a false memory) (Loftus, 1997: 71). Specifically, subjects were made to believe – to believe that they remember – getting lost at the local mall when they never did (Loftus, 1997: 71). In these experiments, family members provided the evidence indicating that research subjects must be creating a false memory, testifying that no such event had occurred (Loftus, 1997: 71).

Misinformation with respect to the reliability of memory is the observed phenomenon whereby research subjects

[43] See Blank and Dieguez (2009) for a detailed discussion concerning the nature, or current understanding of, out-of-body and near-death experiences.

have been shown to remember past events differently when presented with misleading information (Loftus, 2005: 361). For example, subjects were shown a video where a car stopped at a stop sign, subsequently driving away. After this initial event, subjects were fed misinformation – that the sign had really been a stop-and-go. Surprisingly, a statistically significant number of subjects' memories were affected such that they falsely recalled the sign really been a stop-and-go (Loftus, Miller, & Burns, 1978).

On a different note, humans don't appear to be particularly good at divining the true reasons for their behaviour – i.e. beliefs, attitudes, motivations and so forth.. For example, simple contextual cues – the temperature of a drink, or the subliminal presentation of words of differential affect – can lead to measurably different behaviour on the part of the subject "stimulated" by these cues (Williams & Bargh, 2008; Mohan, Mahmood, Wong, Agrawal, Elgendi *et al.,* 2016). As Childers notes:

> Since so many of the causal factors guiding our behaviour operate outside the lamplight of consciousness, our understanding of our own motivations inevitably requires a rational reconstruction (Childers, 2011: 197).

3. Human reasoning: biases, illusions, errors, or fallacies

In determining whether, and to which degree, people's perceptual judgments are accurate, the objective standard is provided by science (Nisbett & Ross, 1980). As discussed, human sensory perception doesn't measure up; eyes can misjudge lengths and scale, ears can fail to measure the distance of sounds appropriately, the nose can perceive the same stimulus as different, or different stimuli as the same, and taste really does seem to have a problem with the colour blue.

When human reasoning – or upstream inferential practice more generally – is subjected to similar scientific treatment, it too fails to represent or judge empirical reality reliably.[44] In other words, when studied in carefully designed experimental contexts, human reasoning is no less immune from error than human sensory perception. Kahneman and Tversky make the point as follows:

> In making predictions and judgments under uncertainty (in reasoning in contexts of imperfect information), people do not appear to follow the calculus of chance or the statistical theory of prediction. Instead, they rely on a limited number of heuristics which sometimes yield reasonable judgments and sometimes lead to severe and systematic errors (Kahneman & Tversky, 1973: 237).

[44] As evidenced by various scientific studies. This is not in contradiction or inconsistent with the naturalist's claims – in Chapter 2 – that human cognition would likely be reliable on naturalism and evolution. The latter reliability refers to the need for humans to have been reliable with respect to features of reality crucial to their survival and reproductive success. For instance, within the context in which they evolved, being right about predators and prey would have been important, while knowing anything about the molecular structure of water or the nature of scientific evidence would have been irrelevant.

And:

> The presence of an error of judgment is demonstrated by comparing people's responses either with an established fact… (or) with an accepted rule of arithmetic, logic, or statistics (Kahneman, Slovic, & Tversky, 1982: 493).

Errors of judgment, biases, or reasoning fallacies are (supposedly) legion (see page 73 and 74 for objections). These include the conjunction fallacy (sometimes referred to as the 'Linda problem'), base-rate neglect or the base-rate fallacy, overconfidence bias, simple logical reasoning errors, and the hot-hand and gamblers fallacies amongst others (Kahneman *et al*., 1982: 154; Kahneman & Tversky, 1983: 299, 300; Gilovich *et al*., 1985; Kahneman, 2011). Below, a brief discussion of each follows.

3.1 The conjunction fallacy (the 'Linda' problem)

Consider the following experiment. Experimental subjects are given a number of biographical details about a person called 'Linda':

Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in antinuclear demonstrations (Kahneman & Tversky, 1983).

Subjects were then asked 'Which is more probable?':

(A) Linda is a bank teller.

(B) Linda is a bank teller and is active in the feminist movement.

(C) In the original experiment, 142 of 167 (85%) research subjects answered B (Linda is a bank teller *and* is active in the feminist movement).

But, according to the rules of the probability calculus, the probability of a conjunction – P(XY) for example – *must* be lower than, or at most equal to, the probability of P(X) or P(Y). In the Linda case, B is the conjunction of A (i.e. Linda is a bank teller) *and* more (i.e. B – Linda is active in the feminist movement), and hence it would be irrational for subjects to give the correct answer as B. The probability of B *must* be lower than or equal to the probability of A. Those who answered B were said to have committed the conjunction fallacy (Kahneman & Tversky, 1983).

3.2 Base-rate neglect or the base-rate fallacy

Another experiment seemingly showing that human reasoning departs from optimal rationality is what's called the base-rate fallacy or base rate neglect (Kahneman *et al*., 1982: 154). This fallacy occurs when inferences are made with respect to the probability of some event E occurring without considering the rate or frequency (i.e.

base rate) with which the event (E) *has* occurred. More specifically, and formally, if the historical probability or frequency of event E is known – P(E) is given – then *this* probability shouldn't be ignored when making a judgment with regards to the probability of a further event of type E occurring. In other words, suppose E* is a possible *future* event of type E, then the *historical* relative frequency or base rate of events of type E shouldn't be ignored in determining the probability of event E* occurring. In determining the probability of P(E* | E), P(E) should form part of the equation. But people don't seem to care much for base rates. A classic example involved sixty students and members of staff at Harvard Medical School (Cassells, Schoenberger, & Grayboys, 1978). The experimental subjects were asked the following question:

> [Suppose] a test to detect a disease whose prevalence is 1/1000 has a false positive rate of 5%, what is the chance that a person found to have a positive result actually has the disease, assuming you know nothing about the person's symptoms... (Casscells *et al*., 1978: 999)?

The correct answer – employing Bayesian probabilistic inference – would be that the chances that the person has the disease, given that she tested positive, is one over fifty-one, or slightly less than two percent. However, despite (presumably) being sufficiently informed about statistical inferential procedures, about half of the experimental subjects answered that it was ninety five percent probable that she was infected given the positive test result, whereas only eighteen percent gave the statistically correct answer (Casscells *et al*., 1978: 999 - 1001; Gigerenzer, 1991: 9).

Consider another (constructed) example. Suppose someone prays for a family member to recover from an incurable cancer – and she does. As a result, assume this person is absolutely convinced that his prayer was the cause of her recovery; God listened, and answered. Would he be rational in concluding that his prayer – and God's resultant intervention – *was* the cause of the recovery? Not necessarily. For, if the rate of recovery from the relevant cancer – without any medically understood cause – is five percent in the population at large (with the family members of both the religious and the irreligious recovering with no statistically significant difference), he wouldn't be rational in concluding that his family member was *certainly* cured by God, even if she *was*. Probability theory dictates that the strength of his belief that his prayer was the cause of the recovery should be tempered by the background information that some people – five percent of the relevant cancer-ridden population – recovers with no *known* cause.

3.3 Overconfidence bias

Overconfidence bias is the experimentally verified finding, or more accurately, the interpretation of experimental results, that people's confidence in the accuracy of their probability judgments is systematically higher than it ought to be (Kahneman, 2011) For example, when research subjects were asked to answer general knowledge questions like 'which city is the most populous? (a) Chicago or (b) Houston, and then asked to rate how confident (in percentage terms) they were that their answers were correct, it was found that subjects overestimated the accuracy of their probability judgments by about twenty-five-percent on average (Lichtenstein, Fischhoff, & Phillips, 1982; Gigerenzer, 1991). For example, when subjects were *certain* that all their answers to a series of

such questions were correct, the actual relative frequency of correct answers was eighty percent. If they were ninety percent sure, the real figure was seventy five percent, and so on (Gigerenzer, 1991).

3.4 The gamblers and hot-hand fallacies

The gambler's fallacy, is the gut feeling, or explicit belief, that a run of bad cards, "crappy" rolls of the dice, or spins of the roulette wheel, are *bound* to be followed by luck of a more amicable kind (Gilovich, Vallone, & Tversky, 1985). In other words, the gamblers fallacy is the intuitive belief on the gamblers part that his luck *must* turn; that chances of him winning must be increasing with each loss. But games of chance are designed such that each event – whether rolls of the die or spins of the roulette wheel – are statistically independent (Gilovich, *et al*., 1985; Rabin & Vayanos, 2010). Simply put, dice don't have memories, neither do roulette wheels; getting a pair of snake eyes or landing on lucky number seven *remains* 1/36, *regardless* of what has happened in the past. Still, Vegas remains a gamblers paradise.

The hot-hand fallacy is the mirror image of the gamblers fallacy. Here, the belief is not that a string of losses increases the chances of winning, but that a string of favourable events is likely to continue (Gilovich *et al*., 1985; Rabin & Vayanos, 2010). It's a fallacy because – when engaged in games of chance – each play of the relevant game is designed to be statistically independent (Gilovich *et al*., 1985). Like the gambler who believes that his luck must eventually turn – that his chances for a win must be increasing with each loss – the hot-hand believes that her good fortune is set to continue, despite the fact that the probability distribution of each roll of the die or spin of the roulette wheel remains the same. A rational epistemic agent wouldn't take part in such games to improve their financial well-being.

3.5 Confirmation bias

Human reasoning with respect to causality (often) misjudges true causality from what are really cases of correlation (Fales, 2002). For example, in identifying what they take to be the true cause of some phenomenon, humans tend to note the instances where A leads to B, but fail to ask the necessary question whether B occurs, or has ever occurred, in the *absence* of A. More generally, research shows that humans are unduly interested in evidence or arguments that are agreeable to their current point of view, while conflicting evidence or arguments are either ignored or devalued; a phenomenon psychologists call confirmation bias (Kahneman, 2011).

3.6 The Wason selection task: When humans dim the lights on Modus Ponens and Modus Tollens

Probably the most famous example of human inferential malpractice is what's referred to in the literature as the Wason selection task (Wason, 1966; Wason, 1968). This task was designed to test participant's prowess with respect to simple deductive reasoning. Specifically, researchers wanted to determine if subjects could distinguish, and apply, the *only* two logically valid deductive reasoning schemas in search of a solution to the problem presented.

In the original experiment subjects were presented with four "cards", on each of which a letter or a number was

printed, both on the front and the back (Wason, 1966; Wason, 1968). For example, if a card had a number printed on the front (odd or even) then it would have a letter (vowel or consonant) printed on the back. Suppose the first card had the letter E on the front, the second the letter K, the third the number 2, and the fourth the number 7. Subjects were then asked which *two* of E, K, 2, or 7, would need to turned over to confirm that an even number was invariably printed on the back when a vowel was printed on the front. The correct answer would be to turn over the E-and 7-cards. Why?

Turning over the E-card would show whether an even number is printed on the back, as it should be if the hypothesis were true. In other words, turning over the E-card would confirm whether 'If a vowel (E for example), then an even number (2 for example)'. If this isn't the case, the hypothesis is false; if E is turned over and there's no even number, the hypothesis is false.

Suppose there *is* an even number on the back. Would the hypothesis be confirmed or falsified? Not yet. To confirm that it holds, one would have to show that there's no card in the four-card line-up for which the logical form 'If an *un*even number, then a vowel' is true. Turning over K wouldn't show this. For the hypothesis doesn't make any claim about finding a specific number (odd or even) on the back of a *consonant* card (like K). K could have an even *or* odd number printed on the back. It wouldn't matter. Logically, turning over K would be an example of denying the antecedent (i.e. not-a-vowel), and claiming that this establishes the falsity of the consequent (i.e. not-an-even-number), which is deductively invalid.

Turning over the 2-card and finding a vowel *or* consonant printed on the back wouldn't be relevant in determining the truth-value of the hypothesis either. How so? The hypothesis makes a claim about what one should find printed on the back of a *vowel* card – an even number – *not* about what one should find on the back of an *even-number* card. The truth of the original hypothesis – 'If a vowel, then an even number' – is consistent with – 'If an even number, then a consonant'.

However, turning over the 7-card, and finding a vowel on the back, *would* prove the hypothesis false. For it says that a vowel on the front *must* have an even number printed on the back; there *cannot* be an uneven number on the back of a vowel-card. For example, if there's an uneven number printed on the front (e.g. '7'), there can't be a vowel printed on the back (e.g. 'E') – i.e. 'If an uneven number, then no vowel'.

The Wason selection task appears to have established that test subjects have few problems in recognizing Modens Ponens – i.e. If P, then Q, Q, therefore P – and applying it (even if not explicitly). However, most choose to turn over the 2-card – or an equivalent in similar Wason-like experimental "setups" – thus invalidly affirming the consequent, while neglecting to implement what would be the logically valid, and possibly sound, Modens Tollens – i.e. If not-Q, then not-P, not-Q, therefore not-P (Wason, 1966; Wason, 1968).

3.7 Rather-safe-than-sorry reasoning

Perhaps the most important example of human reasoning deviating from the norms of rationality – at least for the

purposes of this thesis – is what's referred to in the literature as the Garcia effect (Ramsey, 2002: 24, 25; Stich, 1990: 61 - 63). The Garcia effect is the name given to the bias on the part of creatures to make what statisticians refer to as type 1 errors more often than they ought (Ramsey, 2002). Simply put, it means that whenever there are asymmetric fitness costs in making one type of error vis-à-vis another, evolutionary successful creatures tend to make the less costly error (Haselton & Buss, 2000; Haselton & Nettleton, 2005) For instance, thinking predators present when they're not (a type 1 error) is less evolutionary costly than making the contrary error – thinking they're not when they are (what statisticians call type 2 errors, or false negatives). In other words, evolutionary successful creatures would rather flee the scene when there's *no* danger than fail to get going when there *is*. In layman's terms, evolutionary fit creatures seem to operate by a rather-safe-than-sorry rule even if it means making statistically significant errors, a rule which (sufficiently) reliable thinkers wouldn't likely employ. For they wouldn't – on average – think predators present when they're not, nor misjudge their absence.

Rather-safe-than-sorry reasoning appears a common human inferential malady. Examples include predator avoidance behaviour, male sexual *over*-perception (thinking females more sexually interested than they are), avoiding the "sick" – rather steering clear of those who appear morphologically abnormal for fear that they may be diseased, and so on (Haselton & Galperin, 2011). The common feature – that which seems to make these examples of inferential *biases* – is that humans appear to form beliefs which aren't reliable, and further, aren't reliable in a systematic manner – i.e. *consistently* making the less costly evolutionary error.[45]

3.8 Human cognitive reliability with respect to metaphysics

Finally, and of particular importance, is the fact that human beings have a hard time in coming to grips with abstract reasoning, and even if they *do* master the intricacies of formal reasoning, the evidence suggests that – at least in metaphysical matters – the results or reliability of such reasoning should be considered with circumspection. Why? Because many intellectual luminaries, in centuries past, and at present, haven't been able to agree, and in fact hold to diametrically opposing views with respect to matters metaphysical (Bourget &

---

[45] See McKay & Efferson (2010) for a contrary interpretation of the evidence. They argue that instances of biased behaviour – i.e. male sexual over-perception *etc.* – is not necessarily the result of the relevant subjects holding and acting on biased *belief*. According to them, it could just as well be the effects of creatures *accurately* estimating the risk-reward payoff structure of the relevant scenario and acting in such a way as to maximize the expected reproductive payoff in that context (McKay & Efferson 2010). In other words, the relevant party could be *behaving* in a biased manner without their *beliefs* being biased.

Haselton & Galperin (2011) respond that there are 'compelling empirical reasons' to conclude that systematic errors in *belief* rather than *behaviour* is the more plausible explanation of (some) of the evidence (Haselton & Galperin, 2011: 23). They explain:

> The notion that selection should not bias beliefs is difficult to reconcile with the fact that men appear to overestimate women's interest in all of these varied ways, but especially in self-report measures. Self-reported estimations reflect biased *beliefs* rather than biased actions (Haselton & Galperin, 2011: 24).

Chalmers, 2014). They cannot all be correct, and even if one party *is* correct, the other parties in the relevant discussion are presumably also aiming at discovering the truth. Simply put, how reliable can human cognitive faculties be – at least with respect to these sorts of questions – if many bright lights of the intellect alight on different answers; answers that often contradict one another?

For instance, some argue that there's a god, others that there isn't (Bourget & Chalmers, 2014: 15). Moreover, those who claim that there is a god cannot seem to agree on how many there are. Some are convinced that the Lord thy God is one (Deuteronomy 6: 4). Others are similarly certain that there are legion (Ridgeon, 2003). And still others that there aren't any gods at all (Dawkins, 2006). The fact that humans appear to be unreliable with respect to their knowledge of ultimate reality – *meta*-physics – doesn't imply or preclude that they cannot ever be, or aren't moving towards, a more reliable conception of such fundamentals. For it may be, or at least turn out, that there is *increasing* agreement on matters far removed from experience – at least as far as the empirical structure of reality is concerned; scientific progress being the prime exemplar.

3.9 Objections: Biases that aren't biases, errors that aren't errors, and fallacies that aren't fallacies

A number of researchers – those of the so-called 'ecological rationality' school – have argued that the biases-and-heuristics research establishment has vastly overestimated the degree to which human cognition is fallible (Gigerenzer, 1991; Gigerenzer & Selten, 2001; Cosmides & Tooby, 1994). Many of the cognitive biases or fallacies of reason are in fact no such thing. Their quarrel can be understood as the development of two claims:

One, there's no set of universally agreed upon standards or norms of rational inference – hence, no *single* answer as to whether some phenomenon *is* a bias or not. As Gigerenzer notes:

> Despite the widespread rhetoric of a single "normative theory of prediction," it should be kept in mind that the problem of inductive reasoning still has no universal solution (the "scandal of philosophy"), *but many competing ones*. The controversies between the Fisherians, the Neyman-Pearsonians, and the Bayesians are evidence of this unresolved rivalry (Gigerenzer, 1991: 88, my emphasis).

Two, the working and rationality of the human mind should be evaluated in light of the nature of the contexts (domains) in which they evolved, and with respect to the problems they were likely tasked to solve, not in artificial lab "set-ups" so to speak. Simply put, to pass judgment on the rationality of human reason, one has to evaluate it where it needed to be exercised – in an evolutionary relevant environment, facing problems of adaptation. As Cosmides and Tooby state:

> [T]he ecological school claims that – viewed in the right context (an evolutionary relevant one) – 'the human mind is not worse than rational... but may often be better than rational (Cosmides & Tooby, 1994: 329).

Consider the first claim – that there's no set of universally agreed upon standards or norms of rational inference. Why is this important? The reason is that if there's no *one* answer or set of standards with which one may compare

some example of human judgment under uncertainty, then claiming that this or that example of reasoning *is* an error or bias may not be the last word on whether it in fact *is*. For, according to Gigerenzer:

> What is called in the heuristics and biases literature the "normative theory of probability" or the like is in fact a very narrow kind of neo-Bayesian view that is shared by some theoretical economists and cognitive psychologists, and to a lesser degree by practitioners in business, law, and artificial intelligence. It is *not* shared by proponents of the frequentist view of probability that dominates today's statistics departments, nor by proponents of many other views; it is not even shared by all Bayesians (Gigerenzer, 1991: 87).

Further, Gigerenzer argues that if one reformulates the research questions used to establish the presence of biases and fallacies like overconfidence, base-rate neglect, and the conjunction fallacy, these errors of reasoning disappear (Gigerenzer, 1991). More specifically, he claims that if one assumes that the human mind is an intuitive frequentist – as opposed to a natural Bayesian – and recasts the research questions accordingly, the reasoning errors referred to (largely) dissolves (Gigerenzer, 1991).

Suppose one agrees with Gigerenzer that overconfidence, base-rate neglect, and the conjunction fallacy aren't biases, errors, or fallacies. Does this show that human reason is error-free – an organic machine 'often better than rational'? Not quite. For one may rightly complain that showing three – albeit renowned – examples of cognitive error or bias to be illusory doesn't establish that *all* suggested errors of human reasoning aren't.

Two salient counter-examples appear to be the gamblers and hot-hand fallacies (Boudry *et al*., 2015). For instance, it's not a stretch to think that at least *some* gamblers know that their luck doesn't *have* to turn – that their belief that their luck has to turn is based more on a hope and a prayer than a rational consideration of the facts (Boudry *et al*., 2015). In other words, it's not unreasonable to suppose that some gamblers know that the house always wins (in the end). Alas, many remain regular participants. And isn't this a clear example of irrational behaviour, even on a frequentist interpretation of probability? It appears so. For even though it may be *evolutionary* rational – a point which will be discussed in section 4 – an experienced gambler has:

> [A]ll the requisite evidence he needs to conclude that this is not a game (roulette) to which his (evolutionary) pattern detection heuristics will apply. If even a crash course in statistics and a careful inspection of the roulette wheel fail to cure him of his habit of thought, then we may… call his reasoning fallacious (Boudry *et al*., 2015: 531).

3.10 Human cognition: Not perfectly reliable or rational

Briefly, human cognition is far from perfectly reliable or rational. Human sensory perception – in all its different modalities – is subject to illusions, and can be fooled for the purposes of entertainment, profit, or scientific study. Human inferential practices stray from the dictates of optimal rationality. Memory isn't the unbiased recording device one might suppose, and neither does it appear that humans are particularly good at identifying the real reasons informing their behaviour. Finally, it's evident that all but few find coming to grips with the arcana of mathematics or formal reasoning easy, whereas coming to an "understanding" of procreation requires little more

than reaching the age of puberty. The question I turn to next is: 'On which is the relevant evidence better explained, evolutionary naturalism or theistic evolution? On which would the facts be more probable?'[46]

4. Naturalistic explanations of the distribution of human cognitive *un*reliability

Naturalists offering explanations of human cognitive reliability – or in this case, the distribution of human cognitive *un*reliability – generally employ what they take to be three central tenets of evolutionary theory. Firstly, 'evolution is opportunistic: it can only work with the materials at hand' (Fales, 2002: 57). Secondly, it depends for its continual creativity on 'lucky mutations, and... selective pressures that are common and persistent' (Fales, 2002: 57). And thirdly, 'evolution can permit some degree of mal-adaptiveness, especially when associated benefits outweigh costs' (Fales, 2002: 57).

In other words, evolutionary naturalists think that a systematic, sensible explanation of each of the errors, biases, illusions, or fallacies of human cognition – broadly conceived – can be constructed on the basis of the following three claims. One, evolution has neither forethought nor foreknowledge. Two, it (primarily) cares for the reproductive success of creatures. And three, it has to work with the limited resources it has, not with those it wished it did (Fales, 2002: 57; Boudry & Vlerick, 2014: 68, 69). Simply put, evolution is a blind, path-dependent, and resource-constrained process.

4.1 Explaining human sensory-perceptual biases

Consider human sensory-perceptual illusions, such as those that afflict vision (for example).[47] Naturalists argue that visual illusions result – or are made possible – by the fact that perfectly reliable ocular indication wasn't necessary for survival in the context in which human sight evolved. Good enough doesn't imply perfection, and perfection – *tout court* – isn't necessary for reproductive fitness.[48] (Kahneman, 2011: 26, 27, 100). I say '*tout court',* because there's evidence that evolution (often?) *does* find the optimal solution to an environmental challenge. For example, there's evidence that some creatures perform a mathematically optimal search of their

[46] The reader may rightly wonder why so much of what has heretofore being discussed in Section 3 involved the notion of the *rationality* of human reason and not its *reliability*? Was the aim not to consider the evidence with respect to the distribution or contours of human cognitive *reliability*? If these terms aren't synonyms – and they're not – why the apparent shift in focus? How will the discussion of the evidence concerning the distribution or shape of human rationality shed light – if any – on the facts concerning their reliability and aid in their explanation?

The short answer is that there's a crucial difference between rationality at the *evolutionary* level and rationality at the level of the (modern) *individual* (Boudry*, et al*., 2015). This difference will prove key in explaining why evolution (sometimes) sacrifices the *reliability* of individuals' beliefs – and thus their rationality – while remaining "rational" in pursuit of its ultimate goal – i.e. the reproductive success of creatures. In other words, evolution can be "rational" at *its* level – which should be expected given its primary interest – while on occasion sacrificing the cognitive reliability and rationality of the individual. The development of this answer follows in Section 4.2.

[47] Similar points apply to the other sensory modalities.

[48] As shown by the continued existence and reproductive success of creatures whose sensory modalities have demonstrable "blind spots" – i.e. Homo Sapiens (Kahneman, 2011).

environment in search of scarce resources when there are no environmental clues as to where these resources might be (Sims, Southall, Humphries, Hays, Bradshaw *et al*., 2008). Also, research suggests that the human brain's network design is optimal or near-optimal in that it maximizes the efficiency with which signals can travel within the brain while minimizing the number of connections required to do so (Gulyás, Bíró, Kőrösi, Rétvári, & Krioukov, 2015). As Orgel's second rule states: 'Evolution is cleverer than you are' (Dunitz & Joyce, 2013: 11).[49]

Still, the evidence does indicate that human sensory perception can be fooled when confronted by environments sufficiently *foreign* to their native environments, even if their design is very reliable in the contexts in which they evolved (Haselton *et al*., 2005). The evolutionary naturalist shouldn't find this surprising, for evolution is blind with respect to the future. It searches the fitness landscape in the dark and its success is measured in hindsight (Fales, 2002). Hence, if the future proves to be much different than the past, or more accurately, much different *too quickly*, it shouldn't be surprising to see a blind-search algorithm like natural selection fail to answer nature's questions sooner rather than later.

What should be gleaned from the above discussion is this: it wouldn't be surprising – on evolutionary naturalism – to find a modicum of unreliability even in creatures well-adapted to their environments. For given the constrained nature of naturalistic evolution, it's not surprising that evolution hasn't toed the line of perfection. For in its unguided meandering through fitness space, it's highly unlikely that each step into the unknown would have matched the contours of perfection, as in fact it didn't, or that it will do so in future. Although the continued evolutionary survival of creatures would certainly require remaining on trajectories satisfying the minimal conditions congenial to reproductive success, this by no means implies that these were also the best possible routes available. It seems to have done the best it could, and the best appears to have been good enough (for many).

Hence, the profile or facts with respect to the unreliability of human sensory perception doesn't appear to pose any real problems for the evolutionary naturalist. Even though she may not be able to fill in the *details* of why some *particular* sensory perceptual illusion occurs (as yet), there are good reasons to think that the framework of evolutionary theory has the conceptual resources to locate these data points somewhere within its explanatory domain, and to expect further details to be uncovered in due course.

4.2 Explaining human reasoning errors

For the evolutionary naturalist, the facts with respect to human sensory perceptual reliability may be unproblematic, even expected, but what is she to make of systematic *inferential* errors like the Garcia effect – i.e. the tendency of creatures to engage in rather-safe-than-sorry reasoning? Or the fact that (most) humans have great difficulty in mastering disciplines centred on formal reasoning, while the average pubescent takes to, or understands, the ins-and-outs of human reproduction with ease? Further, how is she to explain the average

[49] For the reader who might be interested, Orgel's first rule states that: 'Whenever a spontaneous process is too slow or too inefficient a protein will evolve to speed it up or make it more efficient' (Dunitz & Joyce, 2013: 11).

person's tendency to seek out evidence that confirms *currently* held beliefs, while ignoring or devaluing disconfirming evidence? Finally, why do those who have dedicated their lives to reasoning proper disagree on matters metaphysical to the extent that they do?

The scope and depth of the literature aiming to account for human reasoning error is immense. Hence – given the evolutionary focus of this thesis – the discussion to follow will focus on those models that explicitly aim to account for the facts of human cognitive error from an 'adaptationist' perspective (Haselton *et al*., 2005).[50] In other words, models whose point of explanatory departure is that the contexts in which human cognition developed, and the likely problems human thinkers faced in those contexts, are essential factors in explaining why and how humans reason in the manner that they do (Haselton *et al*., 2005).

In general, evolutionary naturalists seed their explanations of human reasoning error in familiar ground; that evolution is unguided, primarily cares for survival, and has to make do with what it has, not with what it wished it did. It's a blind and largely carefree explorer of a circumscribed fitness landscape (Fales, 2002). Given that evolution is unguided – that it lacks foresight – it's no surprise that humans have to make judgments under *uncertainty* – i.e. that they have to form beliefs and engage in behaviour with imperfect information. In general, it's therefore unsurprising – trivial even – to think that human judgments should err to some degree, as the empirical evidence confirms. But, more interestingly, humans not only make reasoning errors, they appear to do so *non*-randomly – i.e. their errors in reasoning aren't randomly scattered about the mean of optimal reasoning, but appear to drift *systematically* from such norms (Kahneman & Tversky, 1973; Kahneman *et al*., 1982). What explains the non-random nature of these errors?

Firstly, the fact is that evolution likely didn't have the resources available to build brains that could map or track the empirical features of reality perfectly. And, secondly, even if it did, it may have "elected" not do so, perhaps "finding" that trading off (some) reliability for other capacities or ways of thinking proved a more optimal strategy in maximizing the chances of creatures' reproductive success (Haselton *et al*., 2005). Simply put – given the nature of evolution – it may have been more rational for *evolution* to build brains that could be accused of *individual* irrationality and unreliability in the service of *its* interests – i.e. evolutionary fitness (Boudry *et al*., 2015). Moreover, if evolutionary rationality likely required trading off individual cognitive rationality and reliability in a systematic or predictable way, it could explain why individual human cognition suffers from the cognitive biases, errors, illusions, or fallacies discussed, or at least some of them (Haselton *et al*., 2005).

Thirdly, and finally, many of the features of the modern world are the products not of the glacial process of biological evolutional, but its speedy cultural cousin. The modern world is a vastly different and more complex place than those in which humans evolved. Hence, it wouldn't be surprising if the innate (modular) cognitive equipment evolutionary psychologists think humans come equipped with don't have as tight a grip on the

[50] More specifically, only two or three such 'adaptationist' models of human reasoning will be discussed here. The goal is to highlight the explanatory job that these sorts of models can do. For an in-depth discussion, see Pinker (1997); Gigerenzer & Selten (2001); Mercier and Sperber (2011).

empirical features of the contemporary world. And hence, that their 'fast-and-frugal' solutions to the questions the modern world asks aren't as accurate as they once were (Haselton *et al*., 2005).

In summary, resource constraints likely meant that evolution had to trade off cognitive reliability for other cognitive and evolutionary goods. Further, given that it aims foremost for the survival of creatures, not the truth of their beliefs, circumstances of a certain nature may have required sacrificing individual's cognitive rationality and reliability for the greater good of reproductive success. Finally, cultural evolution has changed the nature of the world in which humans have to make their living quicker than natural selection has had time to respond. It would therefore be unsurprising to find that the adaptive tools biological evolution has bestowed on human's prehistoric kin prove less effective in navigating the modern world.

In what follows, you will find a brief discussion of some of the more specific theories proposed to make sense of human cognitive biases and reasoning errors. As noted earlier, these models approach the relevant explanatory challenge from an adaptationist perspective.

4.2.1 Error Management Theory: Explaining rather-safe-than-sorry reasoning

An example of the systematic development of the idea that *evolutionary* rationality can lead to and explain the fact that *individual* human cognition (often) proves less than optimally reliable or rational is error management theory (EMT). According to Haselton and Galperin:

> Error management theory predicts that biases will evolve in human judgments and decisions whenever the following criteria are met: (a) the decision had recurrent impacts on (evolutionary) fitness (reproductive success), (b) the decision is based on uncertain information, and (c) the costs of false-positive and false-negative errors associated with that decision were recurrently asymmetrical over evolutionary time (Haselton & Galperin, 2011: 4).

Error management theorists claim that their model explains why humans engage in fallacious rather-safe-than-sorry reasoning. How so? Recall that the fallacy of rather-safe-than-sorry reasoning is the phenomenon whereby humans are prone to commit type 1 statistical errors – e.g. thinking a predator present when it's not – more often than they ought.[51] And when would one expect them (whose faculties are the product of evolution*)* to do so – to make such errors? When the costs and benefits – in terms of *evolutionary* fitness – of making type 1 and 2 errors were *consistently* different in the contexts in which human cognition evolved (Haselton & Galperin, 2011: 4).

[51] The 'fallacy' in rather-safe-than-sorry reasoning is determined by Bayesian norms of inference – the gold standard, or theoretically best – one could possibility do in making judgments under uncertainty (Efferson & McKay, 2010). Hence, a cognitive agent would be making reasoning errors, or be biased in its judgments, were its beliefs to:

> [D]epart systematically from those of an agent with Bayesian beliefs. Such a cognitively biased individual does not have beliefs that are theoretically optimal given the available information. Moreover, the individual's beliefs depart from the theoretical optimum in a systematic, rather than a random, fashion (Efferson & McKay, 2010: 6).

Hence, to show that error management theory can explain putative examples of rather-safe-than-sorry reasoning; male sexual over-perception, (overestimating) the presence of danger, the (irrational) tendency of avoiding the sick, and the persistently prejudicial us-versus-them attitude people display – amongst others – one would need to show that:

> (a) The decision had recurrent impacts on fitness (reproductive success).
>
> (b) The decision is, or was, based on uncertain information.
>
> (c) The costs of false-positive and false-negative errors associated with that decision were recurrently asymmetrical over evolutionary time (Haselton & Galperin, 2011: 4).

But to show that 'the decisions had recurrent impacts on fitness', one would first need to show that humans likely faced such situations, or likely engaged in such behaviours. But this seems very likely true for each condition stipulated.

With respect to condition (a) – the decision had recurrent impacts on fitness – it's clear, or at least highly likely, that humans have consistently faced similar social and non-social environmental dangers throughout their evolutionary journey. For example, they would have needed to interact with humans of unknown social disposition with care, manage contexts involving predators sufficiently well, and navigate risky topography with the necessary level of prudence (Haselton *et al*., 2005).

Second, most human decisions are made on the basis of incomplete and/or uncertain information. For there is no guarantee that the future will turn out the way the past has, or the present is.[52] Hence, given that every decision anyone makes is directed at the future, the information on the basis of which any or most decisions are made in the present could be false.[53] Degrees of epistemic uncertainty, and the need to make judgments under such uncertainty, are ubiquitous features of the human condition. Condition (b) is therefore met.

Finally, its highly likely that condition (c) was a consistent feature of the environment within which humans evolved – i.e. that there were asymmetric costs to making one type of error (false positives) compared to another (false negatives). For, in general, predators and dangers of all kinds were very likely a persistent and pernicious presence within the environment in which humans came to be. Hence, from an evolutionary perspective, it would have been better to believe predators present when they weren't, rather than the other way around – i.e. thinking that they were absent when they weren't. Further, and more specifically, rather-safe-than-sorry-reasoning plausibly explains each of male over-perception, the (irrational) tendency of avoiding the sick, and the persistently prejudicial us-versus-them attitude people have been known, and shown, to display.

---

[52] At least with respect to any contingent matters.

[53] Excluding (*a priori*) necessary truths of course.

For instance, the possible rebuff costs accruing to the overzealous male due to a sexually disinterested female would arguably have been less (on average) than missing the possibility of successfully procreating (Haselton *et al*., 2005). Further, it would've been prudent to be wary of strangers, for wrongly "reading" their intent as benign as opposed to malign would have been much costlier than the cost of an initial somewhat cold shoulder compared to the benefit of a warm embrace (Haselton *et al*., 2005). Hence, as far as I can see, error management theory appears to offer a compelling explanatory account of people's tendency to engage in rather-safe-than-sorry reasoning.

4.2.2 Confirmation bias

At first glance, confirmation bias appears to present a serious challenge to those who want to provide a plausible evolutionary account of human cognition. The reason is that it's commonly assumed that the function of human reason – its evolutionary raison d'être – is to 'help individuals achieve greater knowledge and make better decisions' (Mercier & Sperber, 2011: 4). If this is reason's function, then, as Mercier and Sperber point out, 'it's quite puzzling that reason should fall *systematically* short of being impartial, objective, and logical' (Mercier & Sperber, 2011: 4, my emphasis). The problem isn't that reason sometimes fails in its pursuit of these epistemic goals – adaptive faculties don't need to be perfect – but that it appears to be consistently *mis*guided were *these* its goals. And this is surprising. However, Mercier and Sperber argue that there is available a much more evolutionary plausible account of why human reason – in the employ of the individual reasoner – appears to function as it does – biased (and lazy) (Mercier & Sperber, 2011: 10, 11).

4.2.2.1 Reason is for social consumption

As noted, given the nature of evolution – see Chapter 3 – the environmental context within which human reason or cognition developed is crucial in explaining its functional profile. For it to have evolved, and to have remained in the gene pool as a central feature, it had to have earned its adaptive keep. And in a big way, considering the amount of energy and resources evolution invested in its development and maintenance. If so, one should expect *universal* features it displays – e.g. confirmation bias – to be *features* rather than *bugs* (Mercier & Sperber, 2011: 219, my emphasis).[54] In other words, a feature so ingrained in human reasoning practice is likely there for a very good reason. But what could that be; what adaptive function could the my-side bias plausibly serve?

[54] Mercier and Sperber claim that confirmation bias is really a misnomer (Mercier & Sperber, 2011: 218). According to them, people don't struggle when evaluating counterevidence or arguments in which they have little skin in the game. But they do have a hard time remaining impartial when the relevant evidence and arguments are closer to home. As the put it:

> ... [P]eople have no general preference for confirmation. What they find difficult is not looking for counterevidence or arguments in *general*, but only when what is being challenged is their *own* opinion (my emphasis). Reason does not blindly confirm any belief it bears on. Instead, reasoning systematically works to find reasons for our ideas and against others we oppose. It always takes our side. As a result, it is preferable to speak of a *myside bias* rather than of a confirmation bias (Mercier & Sperber, 2011: 218, their emphasis).

In response, Mercier and Sperber argue that reason, and its product – reasons – acquired its considerable adaptive value in a social context. According to them, reason didn't evolve to aid the solitary reasoner in some truth-seeking enterprise, but to enable its employers to *cooperate* (Mercier & Sperber, 2011: 9, 10). And the (primary) means by which reason makes effective cooperation possible is by enabling its users to give 'reasons for justifying [themselves], and... to produce arguments to convince others'. And that it is *this* ability that makes effective cooperation possible (Mercier & Sperber, 2011: 8). As they put it:

> [C]ooperation poses unique problems of coordination and trust... (Firstly), by giving reasons in order to explain and justify themselves, people indicate what motivates and, in their eyes, justifies their ideas and actions. In so doing, they let others know what to expect of them and implicitly indicate what they expect of others. (Secondly), evaluating the reasons of others is uniquely relevant in deciding whom to trust and how to achieve coordination (Mercier & Sperber, 2011: 8).

Further, if cooperation requires overcoming 'unique problems of coordination and trust', and a plausible manner in which this is achieved is by giving reasons to justify oneself and to persuade others, then, according to Mercier and Sperber, one should *expect* people to have a confirmation bias. More accurately, they claim that one should expect people to have a 'my-side bias' (see note 52). As they put it:

> (In a) context in which persuasion is paramount (where the function of reason is to justify oneself and one's action in a social setting so as to aid in group cooperation)... the myside bias makes obvious sense: when defending a point of view, it is a good thing. *It is a feature rather than a bug...* If the functioning of reasoning, when it produces reasons, is to justify one's actions or to convince others, then it [*should*] have a *myside* bias (Mercier & Sperber, 2011: 8, 219, their emphasis).

But without further argument, or at least filling out what is left unsaid, this doesn't really explain why the my-side bias is a good thing or why one should expect it to be an evolutionary feature. Having said this, I don't think arguments justifying the above claim are hard to come by. For instance, it's very plausible to think that people typically – or in large part – act in their self-interest. Further, it's a feature of the human condition that achieving one's ends often requires the help of others. Hence, the rationally self-interested party should find it very useful to be able to persuade others of his or her own of point of view or to convince the relevant audience that his course of action is the one to take. Moreover, given the evolutionary value of group living, there would very likely have been evolutionary pressure to select for these or related abilities.

However, if the individual doesn't have the ability to critically evaluate the arguments of other similarly gifted "persuaders", she may find her own interests defeated as others aim to push their own agendas. And the same would go for every individual in relation to every other. In other words, from every individual's point of view, it would be rational to deceive others if this would further one's goals. But, conversely, it would be damaging if one weren't able to minimize or reduce the likelihood of falling prey to such deception oneself. As Mercer and Sperber put it:

> Humans... differ from other animals in the wealth and breadth of information they share with one another and in the degree to which they rely on this communication... (the) indispensable benefits we get from communication go together with the commensurate vulnerability to misinformation. When we listen to others, what we want is honest information. When we speak to others, it is often in our best interest to mislead them, not necessarily through straightforward lies but by at least distorting, omitting, or exaggerating information so as to better influence them in their opinions and in their actions (Mercier & Sperber, 2011: 8).

Hence, on average, it would pay for individuals to be vigilant when *others* are giving reasons in support of *their* point of view or plan of action, but not be overly concerned with being truthfull in their *own* reason-giving.[55] If Mercier and Sperber are right – that the primary function of reason is to justify oneself and persuade others – then the evidence should show that humans are (on average) better, or rather more interested or careful, in evaluating arguments than providing them. And, indeed, this seems to be the case (Petty & Wegener, 1998; Mercier & Sperber, 2011; Trouche, Sander, & Mercier, 2014).

4.3 The Hot-hand and Gambler's Fallacies.

The hot-hand and gambler's fallacies are the twin names given to the observed phenomena where people believe – or at least behave – as if (some) events within their environmental context are statistically-dependent when they aren't. When people believe that the relevant events are positively correlated – that one sort of event (e.g. "heads" in tossing a fair coin) should more often than not be followed by another of its kind (another "heads") – it's called the hot-hand fallacy. The gambler's fallacy is the converse. It's when people believe that a string of one kind of event – e.g. heads – is more likely than not to be followed by an event of a different sort – e.g. tails (assuming the relevant events – heads and tails – are equiprobable).

Why would people consistently make such mistakes? What plausible adaptive ends could such cognitive practices have served? Nature is a place where events are often reliably positively or negatively correlated – i.e. events in nature are more or less patterned or regular. If evolutionary survival is a function of matching one's beliefs and behaviours to the features of nature, it strongly suggests that evolution would have selected for creatures that were sufficiently adept at pattern recognition. Moreover, creatures' chances of survival is likely an increasing function of their pattern-recognition abilities – i.e. the more attuned they were at discerning nature's regularities the greater their chances of reproductive success would likely have been.

For example, it would likely be better (on average) for a hunting party to infer that returning to an area that had previously proven successful would prove successful again – *even if that's not always the case.* More generally, it would be evolutionary rational to *over*-infer the degree of positive correlation in a given situation *if* such a

[55] Although I think the degree of concern an individual would likely have with respect to his own reason-giving would be a function of his estimate of the epistemic vigilance of the "crowd" he is addressing. When persuading the gullible, almost any reason will likely do. But this wouldn't be the case when addressing a more informed and critically inclined audience. The effective "persuader" will know or have learnt how to read his or her crowd.

correlation holds more often than not. The occasions on which it works will pay for the few times that it doesn't. The same will hold for individuals or groups of individuals that over-infer negative correlations – e.g. rather-safe-than sorry reasoning. When such a negative relationship holds more often than not, even if its weaker than people believe it to be, it would likely be evolutionary rational for them to believe it with the confidence that they do.

However, there are dangers in making such over-inferences. For people may start seeing all sorts of patterns or correlations where there aren't any, as they in fact do. And these inferential over-indulgences may prove costly. Still, it's unlikely that evolution would allow too many costly or irrelevant correlations to be acted on (even if many are believed). It cares for creatures' survival, and there won't be any survivors if they don't see and act on causally relevant patterns often enough.

Briefly, in a world with exploitable regularities, where there are (sufficiently large) asymmetric fitness costs to making one type of error compared to another, it's likely that creatures, if they err, will have been selected for to do so in a systematic or patterned way. The reason is that it's probably more evolutionary costly (on average) for individuals to fail to recognize reliable and useful patterns than to think that there are such patterns when there aren't (any). It would thus have been evolutionary rational – albeit at the expense of the individual's rationality – for evolution to have selected for creatures that systematically make the less costly error. However, in the modern world, environments have been created to take advantage of human's natural tendency to such over-inference. And in *this* world, over-inference is often *not* the less costly.

4.4 Evolutionary naturalism on human cognitive reliability with respect to metaphysics

Experts in metaphysics disagree on what the fundamental nature of reality is (Bourget & Chalmers, 2014: 14 - 16). For example, philosophers disagree on whether the substantial nature of reality is singular (substantially physical *or* mental), or plural (substantially physical *and* mental). Or whether there is one substance – the physical – but more than one type of *property* (the mental and the physical), or even whether there are substances at all (Bourget & Chalmers, 2014: 14 - 16; Ladyman & Ross, 2007). For purposes of this paper, the most important disagreement is whether God exists or not. A significant minority of professional philosophers – (27.2%) – think that some form of theism or something other than atheism is true, most – (72.8%) – don't (Bourget & Chalmers, 2014: 15). This data point suggests that humans – at least philosophers – find it hard to come to reliable terms with respect to the question of God's existence.[56]

[56] It would be interesting to know what the percentage split is between theists and non-theists whose research focus is metaphysics. For it could be argued that this data would be more representative of the merits of the arguments either way, as (many) of the philosophers whose research interests lie elsewhere may not have spent the time (apart from some core readings) on the question of whether God exists or not.

On the other hand, it could also be that those with a theistic bent are more likely to study metaphysics as it's an area of study where naturalism is not clearly the only (plausible) game in town. However, even if, on inspection, the split reverses, such that two-out-of-three (2/3) metaphysicians are theist versus non-theist, the ratios involved would still indicate that there is significant disagreement among experts in such matters. And even if there is an overwhelming majority either way, history shows that the expert majority have (often) being mistaken – e.g. that the earth orbits the sun or the Aristotelian notion that bodies are naturally at rest.

Simply put, the fact that many experts are convinced of what appear to be metaphysically irreconcilable views indicates that humans aren't particularly good at uncovering nature's metaphysical secrets. And, if one or the other party to the God question is right – which is presumably the case – they certainly haven't convinced their fellows of the errors of their views, assuming their fellows really *do* want to know. Should the evolutionary naturalist find these fundamental metaphysical disagreements surprising? Should she be worried?

4.4.1 Human cognitive reliability with respect to metaphysics on *biological* evolution

The evolutionary naturalist shouldn't find it surprising that humans don't seem reliable in their metaphysical musings. For evolutionary success requires only that creatures be able to reliably identify, track, and act on the evolutionary *relevant* features of their environment. And, as I will show, this doesn't require people to know anything about nature's substance.

The features relevant to reproductive success are *empirical* and *structural*. They are the sights, sounds, smells, tastes, and general "feelings" (proprioceptive and haptic) – and the correlations, patterns, or relations that hold between them. From an evolutionary perspective, it doesn't matter whether humans are reliable metaphysicians; whether they know what reality *ultimately* consists in. It requires only that they are sufficiently competent at tracking, connecting, or correlating nature's empirical states of affairs – i.e. that they are good enough at discerning or inferring the evolutionary relevant dynamics of the world as it is *presented* or *appears*.

More specifically, suppose that reality consisted only of a physical or mental substance. If the world were such, creatures and their environments would by definition be either substantially physical or mental only. But, if so, how would natural selection make its "selections"? It cannot select for substance, for by design everything is of the same substance. And if everything is substantially the same, nothing is substantially different – i.e. natural selection can only make selections if there are different things to select from. Selection would only be possible in a substantially undifferentiated world if that substance is differentially arranged or structured. Hence, what matters for creatures' differential reproductive success is their interaction with the *structure* of reality, not its substance. For *what* it is makes no difference to *how* it is, only *that* it is.[57]

But suppose that reality is not only physical or mental – physical or mental properties and substances – but some combination of these. Would the substantial nature of such a world make an evolutionary difference? No, for a creature's reproductive success is solely a function of how a creatures structural features – it's morphology – interacts with nature's structure. So *even* if creatures were somehow substantially mental and physical, it wouldn't

Finally, it would have been insightful to have had some time-series data; to see how the philosophical landscape has changed over time.

[57] It wouldn't matter what substance grounds a particular structural feature. For example, whether it is mental substance M or physical substance P that realizes structural feature S is irrelevant, only that it is *S* that is realized.

matter. For their survival would depend on how these substances are *arranged*, not that they *are* mental and physical.[58]

In summary; the evolutionary naturalist shouldn't find it surprising that natural selection appears not to have selected for reliable cognition with respect to metaphysics (or at least those areas not amenable to a structural or empirically supported analysis). Nature's substance has no bearing on biological fitness. Hence there wouldn't have been selective pressure to acquire such knowledge. But even if the substantial nature of reality made or makes an evolutionary difference, it would be on account of its structure, not its substance. For in a world made exclusively of the physical or the mental, *everything* is physical or mental. Hence it's only how the physical or the mental is *arranged* – assuming equivalent natural laws hold with respect to both – that could be selected for, at least insofar these structural properties have different physical effects and these effects are within the power of natural selection to select for.

The same holds for a world in which reality is physical *and* mental – i.e. consists only in mental and physical substances, properties, or combinations of these. For its only how the mental and physical are combined or related that may possibly be selected for. That they *are* substantially physical and mental cannot. However, these arguments show only that *biological* evolution likely wouldn't have resulted in humans having reliable knowledge with respect to the substantial nature of reality. Plantinga concurs:

> … [E]ven if we thought it likely… that evolution would select for reliable cognitive faculties, this would be so only for those faculties producing beliefs relevant to survival and reproduction. It would not hold, for example, for the mechanisms producing beliefs involved in a logic or mathematics or set theory course (Plantinga, 1993: 233).

Still, even if biological evolution likely didn't gift humans with reliable metaphysical knowledge, it doesn't imply that they couldn't have gained such knowledge elsewhere or otherwise.

4.4.2 Human cognitive reliability with respect to metaphysics on *cultural* evolution

As noted, humans appear biologically ill-equipped at divining the truth with respect to the substantial nature of reality. However, perhaps the human collective could do better? If the empirical success of humans working together as scientists is anything metaphysical to go by, then it's not unreasonable to think that this will be the case.

[58] Naturally, it could be that mental and physical substances – as substances – don't share the same possibilities or constraints with respect to how they can be structured or combined. Hence, the structural arrangement that is selected for *would* ultimately depend on these substantial constraints. But what counts for evolutionary success, and what natural selection selects for – and can select for – is how these substances are structured, not whether they can be structured in this or that way. In short, natural selection takes the metaphysics *as it is* – whatever it is – and selects for their structural arrangements only.

However, it's appears that our current best scientific theories – and those that have come before – haven't been able to settle the question with respect to what their metaphysical status is – i.e. what they are *ultimately* about (Bourget & Chalmers, 2014). If this is the case, then it's not clear that science – *qua* science – will be able settle such ontological matters, despite its extraordinary empirical achievements. Hence, the reliability of human scientific claims may not justifiably transfer to their metaphysical ones.

4.4.2.1 Human cognitive reliability, the most successful empirical game in town, and metaphysics

Although made possible by biological evolution – the development of science has been far too accelerated to have been the result of the glacial process of natural selection. Its origin likely lies in what is commonly referred to as cultural evolution – which includes, but is probably not exhausted by, the human capacity to communicate in a number of different ways (ostensive, spoken, written), the ability to learn by imitation, and to pass on knowledge inter-generationally (Childers, 2011; cf. Vlerick, 2012; Morin, 2016). If science allows humans to know reliably, what is its subject matter? What is it that humans have (increasingly) reliable knowledge *about*?

Doing justice to this question – the truth about *what* science is closing in on – is beyond the scope of this paper, but a few remarks should prove informative.[59] The empirical success of modern science is undeniable. General relativity and quantum mechanics – the foremost contemporary theories in physics – have passed every empirical test scientists have thus far managed to construct (Briggs, Butterfield, & Zeilinger, 2013). Moreover, they are arguably improvements over their predecessors, not only mimicking their empirical success, and correcting their mistakes, but also providing scientists with a rich set of conceptual resources to pursue a more fine-grained investigation of nature's ways (Psillos, 2018). But are scientific theories true or approximately true accounts of what reality *is*, or simply empirically adequate accounts of how reality *presents* itself?'

Suffice to say that this question has been – and is – the subject of an ongoing voluminous and lively debate (Chakravartty, 2017). For current purposes, its sufficient to know that experts disagree (Bourget & Chalmers, 2014). Scientific realists claim that science is converging on what reality *is*, while anti-realists maintain that science isn't an exercise in divining *the* truth, but only an increasingly reliable source of knowledge with respect to its empirical manifestations (Chakravartty, 2017).[60] Moreover, there are also those who claim that this question is ultimately unresolvable or have serious doubts that it can be (Chakravartty, 2017).

But what do these disagreements show? To my mind, it strongly suggests that humans, even as a collective, aren't reliable metaphysicians. For if the most reliable and supremely successful epistemic activity humans engage in doesn't lead most reasonable and genuinely interested truth-seekers to reliable metaphysical knowledge, I don't know what will. For example, every expert agrees on the empirical results of quantum physics; that it enjoys a stunning degree of statistically significant support. But, there are at least two diametrically *opposed* views on what

[59] See Chakravartty (2017) for an in-depth discussion and a comprehensive list of resources for further reading.

[60] Survey data shows that 75.1 % of respondents – professional philosophers – identify as scientific realists, 11.6% as anti-realists, and 13.3% as 'other' (Bourget & Chalmers, 2014: 15).

it ultimately is that quantum physics is providing such accurate predictions about – e.g. the 'Copenhagen' and 'Everettian many-worlds' hypotheses (Myrvold, 2018). But both cannot be right. Hence, at least one of the relevant metaphysical positions cannot justifiably be described as reliable.

If this is right, shouldn't it trouble the evolutionary naturalist? Shouldn't the fact that presumably (equally) informed, intelligent, and appropriately motivated persons fundamentally disagree on metaphysical matters be worrying? For if her metaphysical point of departure is science, and our best science cannot reliably tell us which metaphysics is likely true, or approximately true, how justified can she be in claiming that her *naturalism* is true?

I don't think that the evolutionary naturalist should be unduly troubled. It is true that the "Copenhagens" and the "Everettians" fundamentally disagree about the fundamental nature of the quantum world. But this disagreement is arguably best seen, and most justifiably situated, as a *specific* disagreement within a *larger* naturalistically friendly environment. For it is not as if the relevant parties are suggesting that one needs to move outside the general *physical* realm to solve the specific problem of what it is that quantum physics is metaphysically pointing at. As far as I can see, it's a disagreement about what the *specific* nature of the physical is, not a disagreement about whether it is physical in the first place.

Further, the continued empirical success of the working assumption that every physical state, object, structure, or event in the world is explained by another, suggests and supports the idea that nature ultimately consists only of physical kinds (of *some* sort) (Papineau, 2020). If this is true, the naturalist can acknowledge that her *specific* metaphysical claims are likely unreliable, but justifiably maintain that her *general* view that nature is likely ultimately physical is well-supported.

5. Conclusion

Reliable inferential practices, reliable sensory-perception, and reliable memory are all important factors in explaining *why* humans are reproductively successful.' But it doesn't explain why they are often reliably *un*reliable in their perceptual and inferential practices. The evolutionary naturalist claims that these errors or biases can be made sense of when they are seen as evolution's responses to nature's reproductive challenges. More specifically, they claim that it would often have been rational for evolution to sacrifice the individual's rationality – *in the environment within which humans evolved* – to achieve its primary goal – humans' reproductive success. In short, they claim that evolution's 'ecological' or 'adaptive' rationality explains why individuals are saddled with the cognitive biases they are. Hence, the evolutionary naturalist has no reason to fear such cases of systemic cognitive unreliability.

But given evolution's fixation on reproductive success, there would have been little or no selective pressure to be reliable in subject matters far removed from those relevant to survival – e.g. metaphysics. It *would* have been very important for purposes of survival for humans to have been sufficiently reliable in identifying and tracking the empirical features of their reality (socially and otherwise). And this reliability, coupled with human's social nature, *was* likely the necessary ground in which science would eventually flower. However, reliability in

identifying the structural and dynamic features of reality does *not* guarantee or make probable cognitive reliability with respect to its substantial nature. But given that ontological naturalism *is* a theory about nature's substance, the fact that evolution didn't select for human cognitive reliability with respect to metaphysics *should* trouble the naturalist.

In response to such worries, it was argued that humans working as a group could achieve such reliability despite their individual short-comings, the best example being science. But it was also shown that even our most successful epistemic enterprise hasn't resolved fundamental metaphysical disagreements, *even when those disagreements have been about science itself*. Hence, if even the best epistemic tools we have don't lead presumably genuinely interested parties to an agreeable metaphysical consensus, it's difficult to see what would. One line of response discussed was that the specific metaphysical disagreements within science should perhaps be seen as just that, specific. And that there is likely agreement, or good reason to think, that science provides *general* support to a physicalist or broadly naturalistic worldview.

Given all of the above, I think that the naturalist should be cautious when pronouncing on metaphysical matters. Specifically, she should expect that much of her metaphysics is very likely false, even though there is reason to conclude that her *general* naturalistic view or outlook is scientifically supported. Having said this, I think that evolutionary naturalism offers the better explanation of why one should expect humans to be unreliable metaphysicians, *especially* as individuals. For it appears that the theist should expect the individual to be metaphysically *reliable* – especially with respect to the knowledge of God. That the evidence suggests that this isn't the case should therefore make her uncomfortable. A number of theist responses to this challenge, and their explanation of the evidence with respect to human cognitive biases, are discussed in Chapter 4.

CHAPTER 4: THEISTIC EVOLUTION ON HUMAN COGNITIVE RELIABILITY, THE PROBLEM OF EPISTEMIC EVIL, AND THE HIDDENESS OF GOD.

1. Introduction

The evolutionary naturalist believes that evolution is an unguided process – i.e. that it has neither foresight nor forethought (see Chapter 3). Moreover, its path-dependent and resource constrained – i.e. it has to work with what it has, not with what it wished it did (see Chapter 3). Finally, its primary care is for creatures' reproductive success, not the truth-value of their beliefs (see Chapter 3). According to the evolutionary naturalist, these general ideas provide a sturdy foundation, and flexible scaffolding, within which to ground, and on which to support, a compelling explanatory account of the shape or distribution of human cognitive reliability.

The explanatory challenge the traditional Christian theist faces is clear. She needs to do the same, or better. Which means that she must show that the sort of distributional evidence one sees with respect to human cognitive reliability would be no more surprising on theism than on naturalism. The challenge of explaining why God allows humans to reason in systemically biased ways will be referred to as the *epistemic* problem of evil. The fact that humans are unreliable with respect to their knowledge of God – that he exists and that is *He* that exists – will be referred to as the problem of his *hiddenness*.[61]

To my mind, the problem of God's hiddenness is the much more serious of the two, and therefore it will be the main focus of this chapter. For purposes of this thesis, two examples of the latter – the hiddenness of God – will be discussed. The first is what Stephen Maitzen refers to as the problem of the 'uneven distribution of theistic belief around the world' (Maitzen, 2006: 177). The problem the traditional Christian theist confronts is this: 'Why is the distribution of theistic belief so geographically, socially, or culturally clustered? Why is the knowledge of God so 'patchy' when it is *so* important that every individual – no matter his or her culture – knows that God exists, and that it is *He* that exists?

The second hiddenness-of-God problem is that for most of human history – as far as the relevant experts can determine – they didn't worship anything resembling the God of traditional Christian theism. If Christianity is true, this should be surprising. For why would God allow humanity to be as epistemically misguided with respect to him for so long (as appears to have been the case)? In short, on Christian theism, the historical or time-series distribution of humans God-concepts appears to be all wrong (Marsh, 2013).

The epistemic problem of evil will be discussed in Section 2, the problem of God's hiddenness in Section 3 and 4, with closing remarks following in Section 5.

[61] That is, that the *Christian* God exists.

2. The epistemic problem of evil

On traditional Christian theism, God is morally perfect (omnibenevolent), knows every possible truth (omniscient), can bring about any logically possible state of affairs (omnipotent), and he's everywhere (omnipresent) (Taliaferro & Quinn, 1997). On the current account, he also guides the evolutionary process. But then why guide, steer, or allow the evolutionary process to result in humans systematically forming unreliable beliefs of the types discussed – see Chapter 3? Couldn't God have done a better epistemic job? Given his omnipotence, it appears that he could have, and given his goodness, it seems that he would have. Given traditional Christian theism, it's surprising that he didn't.

Firstly, consider God's power. Given his omnipotence – his ability to bring about any logically possible state of affairs – there appears to be no reason to think that he couldn't have made humans more capable epistemic agents. For there's no logical contradiction in claiming such, and there's no argument that I am aware of from inconceivability or impossibility to this conclusion either.

For example, it's imminently plausible to think that God could have created naturally gifted Bayesians, optimally (or near optimally) updating their beliefs under uncertainty, as opposed to, or instead of, the actually existing kind. Were God to have created humans of this superior cognitive constitution, the result would likely have been one of two possibilities.

One, they would be as evolutionary successful as the less epistemically gifted – i.e. achieve the same level of evolutionary success without been as cognitively unreliable. Or, two, they would have been *more* evolutionary successful. How so? Naturally gifted Bayesians would allocate the finite resources at their disposal more efficiently, and being more efficient would likely have resulted in them being more (evolutionary) successful than those not so cognitively competent (all else equal).[62]

For example, by *not* fleeing when they didn't have to, they wouldn't have expended unnecessary energy, energy that could've been more fruitfully employed. In general, by not reasoning in a less than optimal 'rather-safe-than-sorry manner', the average cognitively "enhanced" human would have been more likely to reproduce successfully than the average Homo Sapien. If God could have created a world in which humans had the same or greater chances of reproductive success, but didn't sacrifice as much truth in its pursuit, why didn't he? Given his omnipotence, I find this somewhat surprising. For one of Plantinga's central claims in the evolutionary argument against naturalism is that we resemble God in being 'knowers'. But then why allow the similarity to be unnecessarily diminished by allowing us to be individually irrational in certain contexts when it could have been otherwise? As Fales puts it: ' (We) clearly... do not "reflect" God's nature as a knower very closely, or even remotely as closely as would be possible in creatures God could make' (Fales, 2002: 53).

Secondly, consider God's goodness. Given his perfect moral character, there's good reason to think that he would have desired that humans were more reliable cognitive agents. For with respect to God's capabilities, the fact that

[62] See Selten (2001) for a contrary view.

it *is* maximal is the reason it's judged a perfection – a God or a person *less* cognitively able would be less praiseworthy in comparison. Granted, humans are finite, and thus, trivially, one shouldn't expect their knowledge, or capacity for knowledge, to be perfect. For knowing less doesn't necessarily imply that one should think a finite epistemic agent blameworthy for being such, but merely less praiseworthy in comparison to someone more accomplished.

However, on God's goodness, it's surprising that finite human cognitive agents make *systematic* epistemic mistakes. The reason is that optimally updating beliefs under uncertainty – i.e. being a Bayesian – is consistent with being finite. In fact, it implies epistemic finitude insofar as the judgments involve *uncertainty*. Making judgments about the future based on the right odds doesn't mean that the relevant agent will (always), or even often, be right. Naturally, it *could* happen that such a Bayesian makes judgments that always turn out true, but this is vanishingly unlikely. What all this boils down to is this: God could have created more epistemically praiseworthy cognitive agents (consistent with them being cognitively finite). Simply put, God could have created creatures that resemble him more closely.

Further, and perhaps more theistically troubling, is the (moral) evil that could arguably been prevented, or the moral good that could have been done or accomplished, were humans better epistemic agents. For example, if medical doctors were better at making judgments under uncertainty – e.g. more accurate in their diagnoses – the well-being of many could have been improved. If these kinds of judgments would have led to a (net) increase in the moral well-being of society, it's surprising that God didn't make humans more cognitively able in this respect.

In short, given that the God of traditional Christian monotheism is all-powerful, it certainly appears that he could have created humans that had greater powers of epistemic judgment and thus bore a closer resemblance to him. Given his perfect goodness, one would have expected him to have done so. For as noted, if humans were more epistemically capable, or less prone to systemic error, the net moral benefit to society would likely have been positive. If this is right, then the fact that God didn't create such a world is theistically unexpected and thus in need of explanation.

2.1 The problem of epistemic evil: theists respond

The theist can agree with the evolutionary naturalist that God could have guided the evolutionary process in such a manner that humans were more epistemically capable than they evidently are. Supposing that God could have chosen to create any logically possible universe, its plausible to think that there's at least *one* universe where God guided evolution and the result were humans capable of updating their judgments under uncertainty optimally (or at least better than what they in fact do).

The theist can respond that it's not clear that naturally gifted Bayesians would have been the better "survivors" in an evolutionary world that is relevantly like ours. For it could be the case that the trade-off between epistemic acuity and other epistemic goods like speed of processing is less favourable on Bayesianism than on the cognitive operating system humans have (Selten, 2001: 17). Further, if the odds of coming to know God – the supreme good

on theism – is an increasing function of humans' chances of survival, then it shouldn't be surprising to find that humans err on the safe side in contexts important to survival. Still, as far as I can see, it's reasonable to think that an *omnipotent* being could have created an evolutionary world *unlike* ours but in which naturally gifted Bayesians would have been the better survivors. And hence, not only would they have been more likely to be reliable with respect to the lesser truths relevant to survival in the here-and-now, but also those invaluable to the hereafter.

However, when considering God's goodness – I think there's much more room for plausible theistic disagreement. For God may have perfectly good reasons (consistent with his goodness) for allowing or steering the evolutionary process such that the human cognitive landscape appears as it does (Tooley, 2019). Granted, it does appear true – on a purely *epistemic* basis – that having the ability to form, hold to, and act on more reliable beliefs is better than not. But it's far from clear that having more reliable beliefs would on average be a net moral positive. For there may be competing moral goods that can only be realized to an extent consistent with God's moral perfection if (some) human cognitive reliability is traded in within the relevant contexts.

For example, rightly judging that one's medical condition is such that survival is highly unlikely may drain one's hope that things could perhaps turn out for the better – where a rosier outlook than merited by the facts could in fact be the deciding factor between life and death.[63] The good of knowledge – knowing the true odds – could decrease the good of hope, such that the net effect is negative; a person dying where he could have survived. And perhaps, in the process of hoping that he could survive – despite the odds suggesting that his hope should be tempered or abandoned – the person could do good to those around him. Moreover, his living longer may also afford him the opportunity to develop the sort of character God wishes him to have – i.e. having more time for the good of 'soul-making' (Tooley, 2019). In other words, the person's lack of knowledge, and his unwarranted or irrational hope could have unforeseen ripple effects to the ultimate net moral benefit of himself and others.

Another popular response some Christian theists employ to resolve the tension between God's perfect moral character and the world's imperfect appearance is to appeal to man's ignorance concerning God's reasons for allowing the world to appear as it does (Howard-Snyder & Moser, 2002a). The sceptical theist claims that there's plenty of room in the gap between humans' epistemic finitude and God's goodness to explain why he allows them to make morally relevant epistemic mistakes. Maitzen explains:

> In response to the argument from [epistemic] evil sceptical theism concedes to a-theology that no known theodicy works: none adequately explains – none gives morally sufficient reasons for God's permitting the amount and variety of [epistemic] suffering our world contains. But sceptical theism insists that the failure of all our theodicies is predictable. According to sceptical theism, theists and atheists alike should accept the conditional claim 'If God exists, then God's morally sufficient reasons for permitting [epistemic] suffering may well be outside our ken', given how feeble our minds are when compared to

[63] There are almost certainly cases where people are aware of the true odds of some event occurring, but choose to ignore it for some reason. Here, the assumption is that there are cases where their beliefs are faulty – that they don't know what the true odds are – but act on what they (falsely) believe them to be.

> the divine mind (Maitzen, 2006: 186).[64]

With respect to explaining the data concerning the distribution of epistemic deficiencies in perception, memory, and inferential practices in general, the sceptical theist has a point – perhaps not a very good one – but one that, taken at the limit, holds. Humans are epistemically finite, and even though the collective may in time approach a kind of omniscience, the fact is that no one can be absolutely sure that God doesn't have reasons for allowing the presence of such epistemic deficiencies. Neither has anyone non-contentiously shown there to be a *logical* contradiction between the claims that God is all-good, all-powerful, and the presence of evil (Tooley, 2019).

However, as an explanation of why human cognition is unreliable in the ways it evidently is, its grossly inadequate compared to the explanation the naturalist has to offer (see Chapter 3). For its as uninformative as to be trivial; it's as if, when asked what causes precipitation of any kind, answering that 'God is the cause'. Even if true, all this says is that everything that is not God is not the cause. It doesn't explain or offer reasons how God does it, why he does it, or why there are different types of precipitation, in different places, and at different times.

Briefly, it seems that in some cases, *not* knowing the true state of affairs, and thinking things are rosier than they actually are, may in fact increase the chances that such (good) states of affairs may come to pass. Examples where generally affective goods or virtues have room to flourish – where humans have the ability to grow in virtue, or have the good of freedom to do so – shows that it's not a simple matter whether knowing more rather than less is always, or mostly, a good. Simply put, when the good of knowing more competes with other goods, the net welfare effect could be negative.[65] And thus, it could be more valuable not to know than to know. Finally, as the sceptical theist maintains, God could have sufficient moral reasons for allowing the apparent epistemic evils he does, reasons which human beings aren't privy to. Simply put, God could be balancing the equation of competing goods such that the *net* benefit to humanity is positive and consistent with his moral perfection.

For the sake of argument, suppose that theists are successful in meeting the sort of epistemic challenges raised thus far. But granting this wouldn't close the book on the problem of epistemic evil. For there's at least one challenge – and to my mind the gravest – that remains. And this is the problem of the hiddenness of God – the problem that humans are evidently unreliable in knowing that God exists and that it is He that exists. It's gravity lies in the supreme importance of its subject matter. In fact, the Bible indicates that those who fail to believe in the Christian God and don't respond to this truth appropriately, either out of ignorance or wilfully, will suffer everlasting damnation (Matthew 26; Jude 1; Mark 9; Romans 1). In other words, on Christian theism, its seriously important that people are reliable with respect to the knowledge of God. That they evidently aren't is a significant theistic problem.

[64] In the original, Maitzen refers to evil in general. But by implication, this would apply to epistemic evil as well, since the latter is a subset of the former.

[65] Quantifying over all morally relevant human desires, thoughts, and actions.

3. Theistic evolution on the reliability of the knowledge of God

As noted, humans disagree on whether God exists, and if he does, that it is he that exists. Moreover, this apparent unreliability with respect to the knowledge of God has been millennial in duration and global in scope, involving billions of people. The theist in general, and the traditional Christian theist in particular, should find this surprising, and a pressing problem in need of explanation.

Why should this apparent unreliability with respect to the knowledge of God – the problem of the hiddenness of God – be particularly troubling for the (traditional) Christian? Firstly, if traditional Christianity is true, those who don't know that God exists or that Jesus is his son face problems of a grossly unappealing and eternal nature – i.e. hell (Matthew 26: 41, 43; Jude 1:7; Mark 9: 43). If "finding" God, at least epistemically, is an eternal good (heaven), and not "finding" him an eternal evil (hell), why are so many – most of the world's population in fact – ignorant of these truths? How does one square this with the claim that God is perfectly good and maximally capable?

Secondly, not only is God seemingly hidden to most, God appears to be more or less hidden to those of different cultures and geographies (Maitzen, 2006: 179). For example, if we assume monotheism, why do those of the Middle East appear to be "closer" to knowing the truth about God – they are mostly monotheistic – as opposed to people from Asia, most of whom don't believe that any sort of traditional monotheism is true (Maitzen, 2006: 179)?[66]

Thirdly, and perhaps most puzzling, is why God has been hidden for *most* of human history (Marsh, 2013: 349). In other words, why has God seemingly favoured recent generations with more accurate godly knowledge when compared to the many more that have come before? For the historical evidence indicates that monotheism, of which traditional Christianity is a prime example, is a *modern* conception of God, no more than a few thousand years old (Wright, 2009*)*.[67] Or, more weakly, it appears the belief that there is only one personal God – monotheism proper – is a more recent addition to humanity's conception of the divine (Marsh, 2013: 363).[68] Thus, even if Christians have the true knowledge of God, for most of history, humans haven't. Why?

---

[66] For the sake of argument, the working assumption here is that traditional Christian theism is true, a crucial element of which is the claim that there is *one* God (ignoring for the purposes of this thesis the complexities the doctrine of the trinity presents with regards to this claim) (Deuteronomy, 6: 4; Mark 12: 29).

[67] Due to space and scope constraints, the problem of the hiddenness of God will be considered in the light of traditional Christian theism only. Thus the moniker 'theist' should henceforth be read as 'traditional Christian theist '– i.e. he or she whose central beliefs includes the idea that God exists and that his son Jesus has come to save the world from eternal damnation (John 3: 16 - 18; Mark 16: 16; John 5: 17, 18).

[68] But even if monotheism *is* the primordial conception of God – from which other conceptions of the divine diverged or evolved – a recognizably *Christian* conception of God is a more recent development than animism, ancestor worship, polytheism, or monolatry (Marsh, 2013; Wright, 2009).

3.1 The hiddenness of God: theists respond

In general, the kinds of responses theists offer to the problem of the hiddenness of God – the apparent unreliability of human cognition with respect to the knowledge of God – are three-fold. One, they claim that those who don't know – i.e. the "non-believers" – are always blameworthy for being so unreliable or non-believing. Two, they argue that non-belief or cognitive unreliability of this kind may be blameless, but that God has good reasons for allowing such non-belief or unreliability. Finally, there are a number of theists who consider the hiddenness of God unproblematic (Maitzen, 2006: 180).

3.1.1 God is not hidden

Christian ripostes of the first kind – blaming the unbeliever for his unbelief – argue that there's no one that is ultimately unaware of God existence – there are no honest atheists or agnostics. In other words, everyone *is* ultimately reliable with respect to the knowledge of God – they *know* that God exists – but fail to engage with this truth appropriately. In support of this claim, consider what the apostle Paul's says in Romans Chapter 1:

> For the wrath of God is revealed from heaven against all ungodliness and unrighteousness of men, who by their unrighteousness *suppress* the truth. *For what can be known about God is plain to them, because God has shown it to them*. For his invisible attributes, namely, his eternal power and divine nature, have been clearly perceived, ever since the creation of the world, in the things that have been made. *So they are without excuse.* For although they knew God, they did not honour him as God or give thanks to him… (Romans 1: 18 - 21, my emphasis).

In other words, God isn't hidden, and has never been. For 'his eternal power and divine nature have been clearly perceived… since the creation of the world'. The real problem isn't that God is hidden, but that (most) humans don't *want* to "find him" – they don't want to 'honour him as God or give thanks to him', but intentionally 'suppress the truth'. They are unbelievers because they want to be.

In response, the atheist or agnostic may rightly wonder why much of the project of theistic philosophy of religion has been to provide *reasons* in support of the truth of the proposition that God exists. If the knowledge of God is as 'clearly perceived' as the apostle Paul suggests, why is there a need for *any* supporting arguments? In other words, why do some theists appear to disagree with the apostle, claiming that there *can* be such a thing as blameless non-belief? Aren't they risking the salvation of atheists and agnostics by entertaining the proposition that unbelievers may be blameless in their unbelief despite what the Bible seems to teach? Why make such arguments when the existence of God isn't in need of any argument? If they *know* that the apostle is right, then offering arguments that aim to show that God exists is disingenuous at best.

To my mind, the fact that most people evidently don't know that *Christian* monotheism is the only *true* religion, and haven't for most of human history, militates strongly against the idea that God has morally sufficient reasons for assigning them eternal blame. To make such a charge stick and be consistent with God's goodness, one would

have to assume that most of the world's human inhabitants, past and present, are, and have been, intellectually dishonest. Moreover, not only did they fail to believe, but they rejected an open invitation to enjoy an eternal loving relationship with a perfect Being. That the overwhelming majority of humanity have rejected *such an offer* appears far-fetched. In short, this response to the problem of the hiddenness of God is inadequate.

3.1.2 The hiddenness of God and blameless non-belief

A more popular response to the problem of God's hiddenness accepts that people may be blameless in their non-belief but claims that God has good reasons for allowing this. According to Howard-Snyder and Moser:

1. God hides and thus permits inculpable non-belief (at least in principle) in order to enable people *freely* to love, trust, and obey Him; other-wise, we would be coerced in a matter incompatible with love.

2. God hides and thus permits inculpable non-belief (at least in principle) in order to prevent a human response based on improper motives (such as fear of punishment).

3. God hides and thus permits inculpable non-belief because, if He were not hidden, humans would relate to God and to their knowledge of God in presumptuous ways and the possibility of developing the inner attitudes essential to a proper relationship with Him would be ipso facto ruled out.

4. God hides and thus permits inculpable non-belief because this hiding prompts us to recognize the wretchedness of life on our own, without God, and thereby stimulates us to search for Him contritely and humbly.

5. God hides and thus permits inculpable non-belief because if He made His existence clear enough to prevent inculpable non-belief, then the sense of risk required for a passionate faith would be objectionably reduced.

6. God hides and thus permits inculpable non-belief because if He made His existence clear enough to prevent inculpable non-belief, temptation to doubt His existence would not be possible, religious diversity would be objectionably reduced, and believers would not have as much opportunity to assist others in starting personal relationships with God.

7. Inculpable non-believers are either well-disposed to love God upon believing or they are not. The well-disposed either are responsible for being so disposed or not. If not, God lets them confirm their good disposition through choices in the face of contrary temptations before making Himself known. If so, they are well-disposed for unfitting reasons and He waits for them to confirm their good disposition in a purer source before making Himself known. Inculpable nonbelievers who are not well-disposed to love God upon believing and who are not responsible for failing to be well-disposed

> are given the opportunity by God to change before He makes Himself known (Howard-Snyder & Moser, 2002: 9, 10).

Suppose that the above claims hold. Would the problem of the hiddenness of God be mitigated, undermined, or solved? Arguably not, for there appears to be at least two sorts of (significant) worries that wouldn't be addressed by claims (1) to (7), either individually, or in concert.

3.2 Problem 1: The uneven distribution of theistic belief

The problem of what Maitzen calls 'the *uneven distribution* of theistic belief around the world' is exactly that: 'Why is theistic belief so 'patchy' – so geographically and culturally clustered (Maitzen, 2006: 177, 180)'? As he points out:

> Those who grant the existence of blameless non-belief, and try to explain why God tolerates it, never ask why God tolerates it so unevenly (Maitzen, 2006: 180).

More specifically, he claims that:

> [W]ith perhaps one exception (point 6 above), they – i.e. those who grant the existence of blameless non-belief – fail to address, let alone explain, the demographics of theism… [O]nly response (6) comes close to addressing the geographic disparity of theistic belief, and then only because it broaches the possibility that 'religious diversity would be objectionably reduced' if God were less hidden (Maitzen, 2006: 182).

Grant for the moment that there is a level beyond which 'it would be objectionable to reduce religious diversity'. However, this appears to be in tension with what Christians consider to be one of God's central commandments – the great commission. It runs as follows:

> Then Jesus came to them and said, 'All authority in heaven and on earth has been given to me. Therefore go and make disciples of *all* nations, baptizing them in the name of the Father and of the Son and of the Holy Spirit, teaching them to obey everything I have commanded you…'(Mathew 28: 18 - 20, my emphasis).

In other words, 'go and make disciples of *all* nations', reducing religious diversity, but don't be *too* successful, for then 'religious diversity would be objectionably reduced'. It also seems to clash with the idea that humans' ultimate good consists in entering a loving relationship with the *Christian* God. Which means – if traditional Christianity is true – that the person who wants to enter such a relationship needs to believe that Jesus is the exclusive way to do so; all other faiths would be false insofar as they deny this (John 14: 6; John 17: 2, 3). But this strongly suggests that religious diversity *would not be a good at all*. Yes, those of other faiths would be raising the world's religious diversity, but in believing as they do they would lose the opportunity of entering into a loving relationship with God – the ultimate good. Indeed, their religiously diversity would earn them everlasting damnation (John 3: 18, 36).

Suppose for the sake of argument that the Christian is successful in diffusing these apparent tensions – religious diversity doesn't clash with God's commands nor his wish for everyone to enter a loving relationship with him. Still, does the idea that God allows blameless non-belief in order to ensure that there would be some degree of religious diversity really solve the demographic problem *as it stands*? For explaining the need for a certain degree of global religious diversity is one thing, explaining why it takes the patchy or clustered *form* it does is quite another.

For example, why is ninety five percent of the population of Saudi Arabia theist while the same percentage of Thailand's non-theist (Maitzen, 2006: 183)? [69] Simply put, if religious diversity is such a good, why doesn't it flourish *within* (some) cultures (Maitzen, 2006: 183)? The religious-diversity-argument can explain why there is a certain degree of religious diversity globally. But I'm not convinced that – on its own – it has the explanatory power to account for the "within-culture" lack of such diversity.

For instance, if religious diversity is at a Godly-appropriate level globally, what makes the Thai such a wretched bunch that God allows ninety-five percent of them to be as theistically clueless as they appear to be? For one couldn't plausibly say that religious diversity would be objectionably reduced in Thailand were the ratio of non-theists-to-theists reduced, for in (most) countries the level of non-theists-to-theists is *lower* (Pew Research Center, 2015). In other words, if a much lower level of religious diversity – specifically non-theist-to-theist – in *other* countries isn't objectionable, it shouldn't be in Thailand. Hence, the religious diversity argument – on its own – doesn't explain the sort of patchy religious demographics one sees in the likes of Thailand.

On the other hand, evolutionary naturalists appear to have it easier when needing to explain the patchy or clustered distribution of religious belief in different countries:

> According to these latter (naturalist) explanations, the patchiness of theistic belief has everything to do with the notoriously haphazard play of human culture and politics and nothing to do with God: the messy, uneven data have messy, uneven causes (Maitzen, 2006: 183).

Having said this, I do think there is a more plausible argument available to the theist in response to Maitzen's demographic challenge. And it is due to Max Baker-Hytch (2016).

3.2.1 A more promising theistic response to the problem of the uneven distribution of theistic belief

Max Baker-Hytch (2016) interprets Maitzen's challenge as a probability claim – that the observed uneven distribution of theistic belief is much less probable on theism than naturalism – i.e. that P(uneven distribution of theistic belief | theism) is much less than P(uneven distribution of theistic belief | naturalism). Baker-Hytch believes that this claim is 'significantly less plausible than Maitzen supposes it to be'(Baker-Hytch, 2016: 376). How so?

[69] In fact, the latest Pew research center survey data indicates that the ratio of theists to non-theists in Saudi Arabia is closer to ninety nine to one percent and that of Thailand seven to ninety three percent (Pew Research Center, 2015).

First, he claims that what he refers to as humans' 'mutual epistemic dependence' – i.e. humans' significant dependence on others for acquiring much of their knowledge about the world – explains the observed uneven distribution of theistic belief (Baker-Hytch, 2016: 376). The manner in which humans come to know things explains why the knowledge of God is so unevenly distributed along socio-cultural and geographic lines (Baker-Hytch, 2016: 376).

Second, he thinks that humans' mutual epistemic dependence would be probable on theism. For he argues that God would likely desire humans to have the opportunity to acquire, practice, and perfect certain goods, goods which would require humans to be (roughly) mutually epistemically dependent in the manner and to the degree that they are (Baker-Hytch, 2016: 380).

Finally, Baker-Hytch claims that the probability of mutual epistemic dependence wouldn't be much lower on theism than naturalism. For, according to him, 'the naturalist too will surely appeal to something like mutual epistemic dependence in order to explain the observed uneven distribution of theistic belief'(Baker-Hytch, 2016: 379).

If the above claims hold, then it would follow – by simple probabilistic aggregation – that the probability of the observed uneven distribution of theistic belief on theism wouldn't be much lower compared to its probability on naturalism. Maitzen's challenge would be met. In what follows, each of Baker-Hytch's claims to this effect will be evaluated.

3.2.2 Mutual epistemic dependence explains the observed uneven distribution of theistic belief

Humans have to rely on one another for much of what they know about the world (Baker-Hytch, 2016: 377). The reason for this is that humans are social in nature, (mostly) live in groups, trust others to varying degrees based on a history of inter-personal interaction, and (they) must be in one place at any given point in time (Baker-Hytch, 2016: 377, 378).

As a result, argues Baker-Hytch, the truth-values of much of what they believe about the world isn't open to (their) direct perceptual verification or inspection (Baker-Hytch, 2016: 378, 379). Hence, if the validity of religious beliefs aren't open to such evaluation or verification – which is likely – then their uneven distribution along socio-cultural and geographic lines shouldn't be surprising (Baker-Hytch, 2016: 379). On the sort of mutual epistemic dependence humans display, it would therefore be probable that theistic belief would be so unevenly distributed as it in fact is (Baker-Hytch, 2016: 379).

3.2.3 Mutual epistemic dependence is likely on theism

Baker-Hytch thinks that were God to create 'intelligent, morally sensitive, (and) free creatures' – e.g. humans – there are likely certain kinds of intellectual goods or virtues that God would want them to have, or have the opportunity to acquire and perfect (Baker-Hytch, 2016: 379, 380). The goods he has in mind are, amongst others, things like 'exercising interpersonal trust versus being invulnerable to deception', or 'sharing responsibility for

one another's acquisition of epistemic goods versus practicing epistemic self-reliance' (Baker-Hytch, 2016: 380, 381). And, according to him, it is the *nature* of these epistemic goods that explains why humans are mutually epistemically dependent to the degree, or in the manner, that they are (Baker-Hytch, 2016: 389).

3.2.3.1 The 'goods': competing (intellectual) virtues

As noted, one example of an epistemic good or virtue Baker-Hytch thinks that God would want human beings to have the opportunity of acquiring, growing in, or perfecting, is that of exercising interpersonal trust versus being invulnerable to deception. As the 'versus' indicates, the good of interpersonal trust is in tension with that of being invulnerable to deception. One cannot really trust another if there's no risk of being deceived; it is them *being* in tension that makes real trust a possibility (Baker-Hytch, 2016: 380). In other words, for something to *be* an act of trust the agent involved would need to be somewhat uncertain about the intentions of the one in whom she is placing her trust. She needs to be vulnerable to deception (Baker-Hytch, 2016: 380). Hence, according to him:

> [H]aving significant opportunities to place interpersonal trust in one another requires our being cognitively limited in roughly the ways in which we in fact are: in particular, it requires that we cannot simply read one another's minds, and more generally, that others can do things in private about which we cannot learn independently of their testimony (Baker-Hytch, 2016: 381).

On the other hand, if humans were significantly less intellectually able, they may not be able to judge whether they are being deceived by their cohorts often enough to preclude a state of affairs too morally miserable given God's perfect moral character (Baker-Hytch, 2016: 380). Trusting without discrimination due to insufficient powers of judgment risks damage to oneself and others, states of affairs which can be avoided if one were more able in making inter-personal judgments.

Simply put, humans need to be "bright" enough to forestall them falling prey to deception too often, while not being so cognitively competent or epistemically aware that there wouldn't be any opportunity for them to practice interpersonal trust. According to Baker-Hytch, humans' evident mutual epistemic dependence' fits the requisite bill – it operates in the equivalent of an epistemic Goldilocks zone. As he puts it:

> [T]he degree to which humans are in fact dependent upon one another's testimony for learning about one another and the world is such as to permit a favourable balance between having significant opportunities to place trust in one another, on the one hand, and yet still being capable of discerning enough about one another via non-testimonial means so as to prevent us from being excessively at risk of deception by one another, on the other hand. (Baker-Hytch, 2016: 381).

Another example of such an intellectual virtue is what Baker Hytch calls 'sharing responsibility for one another's acquisition of epistemic goods versus practicing epistemic self-reliance' (Baker-Hytch, 2016: 381). He explains:

> Sharing responsibility for one another's acquisition of such goods [true belief...knowledge, *etc*.] adds substantially to the range of morally significant free actions that are available to humans, and hence, to

> the range of morally good states of affairs that humans are able to realize... [I]t affords a human person the morally significant choice between depriving others of their knowledge or sharing it with them, between will-fully distorting or suppressing the truth and trying their best to relay it faithfully, between co-operating with others in the search for truth or trying to thwart others in that search, and so on. (Baker-Hytch, 2016: 381, 382)

In other words, on the one hand, what makes such 'sharing... of (intellectual) goods' morally significant – what *makes* it a possible good – is the fact that humans are not *too* epistemically 'self-reliant'. If they were so self-reliant, there wouldn't be any need for sharing such goods, and hence no morally salient choice involved in deciding between sharing or depriving others of such goods. On the other hand, if humans were too epistemically dependent on one another, they wouldn't have the opportunity of practising 'what some epistemologists... regard as a significant good, namely, epistemic self-reliance'.[70]

In short, if humans were significantly more cognitively able, it appears that the 'range of morally significant free actions' and hence the 'range of morally good states of affairs... humans are able to realize' would be reduced (Baker-Hytch, 2016: 381). Conversely, if humans were significantly less intellectually able, they would arguably fall prey to the deceptions of others too often and be unable to enjoy and practice the good of epistemic self-reliance. As Baker-Hytch puts it:

> [T]he way in which humans are actually cognitively constituted is evidently somewhere in between the twin extremes of being so cognitively weak that we are forced to rely upon one another almost continually and being so cognitively well-endowed that we seldom need help from others in order to get at the truth. Our actual cognitive constitution seems to be such as to permit a rather favourable balance between the competing goods of sharing responsibility for one another's intellectual well-being and practising epistemic self-reliance (Baker-Hytch, 2016: 382).

Suppose that Baker-Hytch is right; God would have wanted humans to be able to acquire and practice the sorts of epistemic virtues discussed. And that the nature of these intellectual goods explains why humans are roughly epistemically dependent to the degree, and in the manner, that they are. In other words, assume that he has shown that mutual epistemic dependence is likely on theism. If this is right, it means that he has also shown that the evident uneven distribution of theistic belief around the world is also likely on theism. For if mutual epistemic dependence explains the uneven distribution of theistic belief – which is plausible – and mutual epistemic dependence is likely on theism, then by probability theory, the uneven distribution of theistic belief must also be likely on theism. However, even if it *is* likely on theism, it's possible that it's even *more* likely on naturalism. For as he notes:

[70] One reason for considering epistemic self-reliance a good is that it appears apposite to value intellectual achievements as more praiseworthy were it the product of the work of an individual rather than that of a group. The fact that Einstein, although no doubt relying on the achievements of his predecessors and cohorts, invented or discovered the theory of general relativity *on his own,* arguably makes his achievement that much greater than were it the product of teamwork.

> The naturalist too will surely appeal to something like mutual epistemic dependence in order to explain the uneven distribution of theistic belief. The interesting question... is whether the probability of such mutual epistemic dependence is much higher on naturalism than theism. (Baker Hytch , 2016: 379)?

In other words, the theist doesn't have exclusive intellectual property rights to mutual epistemic dependence. The important question is: 'On which of theism or naturalism is mutual epistemic dependence the better fit?' On which is mutual epistemic dependence more probable? And hence, given that mutual epistemic dependence explains the uneven distribution of theistic belief, on which is the uneven distribution of theistic belief the more probable? One reason to think that mutual epistemic dependence is less likely on theism than naturalism is that it appears to be the more involved of the two explanations. As Baker-Hytch asks:

> [I]s it not worrying that theists need to invoke such a plurality of explanatory mechanisms in trying to account for E [i.e. the observed uneven distribution of theistic belief], whereas the naturalist merely needs to invoke a combination of bio-psychological and cultural factors [i.e. mutual epistemic dependence], factors which the theist (this one at any rate) already grants? Isn't it the case that, even supposing the theist can come up with other explanations which (in conjunction with some- thing like the one offered in the present article) suffice to explain E, the theist is burdened with a considerably less parsimonious and hence intrinsically less probable account of E (Baker-Hytch, 2016: 391)?

He responds that:

> [E]ven supposing that the theistic explanation of the lopsided distribution of theistic belief is more complex than the naturalistic explanation of that phenomenon, it doesn't follow that theism is less simple than naturalism when the evidence under consideration is *widened* to include all the other evidence for which these competing hypotheses seek to account: the existence of a complex, life-friendly physical universe, the conformity of material objects to natural laws, the existence of consciousness, and so on (Baker-Hytch, 2016: 391, his emphasis).

In other words – as Baker-Hytch acknowledges – the explanation he offers in accounting for the uneven global distribution of theistic belief is less parsimonious than its naturalistic rival. It may well be, as Baker-Hytch claims, that theism is the simpler explanation when *everything else* is considered. But the fact of the matter is that everything else is *not* being considered. What needs to be explained is the uneven distribution of theistic belief (around the world). And to that end, naturalism is the better explanation.

Having said that, I think that Baker-Hytch has managed to meet Maitzen's challenge; his is a plausible theistic explanation of why theistic belief is so unevenly distributed globally. For if parsimony is the only clear difference between the theistic and the naturalistic explanations of the relevant evidence, the probability of the patchy distribution of theistic belief on theism wouldn't likely be much lower on theism than it would be on naturalism.

In summary; Baker-Hytch argues that if God wanted to create creatures (humans) to have the opportunity to acquire, practice, and perfect the sorts of intellectual virtues he conceives of, and this requires the sort of mutually

epistemic dependence humans display, then Maitzen's challenge would be met. The probability of the observed uneven distribution of theistic belief wouldn't be much less on theism than on naturalism. For, as he suggests, the naturalist would also likely appeal to something like mutual epistemic dependence to explain why theistic belief is so unevenly distributed around the world. Given that this would be the case, then the crucial difference between the two competing explanations would be one of parsimony; the theistic explanation would be the more complex of the two. But if the theistic explanation isn't *too* involved vis-à-vis its naturalistic rival, then it wouldn't be unreasonable to maintain that probability of the globally observed uneven distribution of theistic belief would be much less on theism than on naturalism.

However, as Baker-Hytch acknowledges, even were his argument successful in defusing worries about the *present* uneven distribution of theistic belief, it doesn't explain why the *historical* distribution of humanities conception of the divine has been so far removed from what the Christian would consider apt (Baker-Hytch, 2016: 393). In other words, there may be a good theistic explanation for why the *cross-sectional* (current) distribution of theistic belief appears as it does, but the same doesn't appear to be the case with respect to the *time-series* (historical) distribution of such belief. This challenge and responses thereto are considered below.

4. Problem 2: Darwin and the problem of natural non-belief

The second sort of worry that Howard-Snyder and Moser's response to the problem of the hiddenness of God – see points 1 to 7 on page 96 – fails to address is what Jason Marsh refers to as 'Darwin and the problem of natural non-belief' (Marsh, 2013: 349). Marsh asks:

> Why, if God designed the human mind, did it take so long for humans to develop theistic concepts and beliefs (Marsh, 2013: 349)?

For, as Marsh notes, evidence with respect to the content of humanity's religious beliefs seems to indicate that for much of their history, (most) humans didn't conceive of God in any recognizable – modern (monotheistic) – way or form. In fact, it appears that:

> [T]he oldest and most widespread form of god concepts is the ancestor spirit or ghost, a type of afterlife belief (Barrett, 2007: 775).

More specifically, it appears – as far as can be determined – that human conceptions of the supernatural evolved or developed (roughly) as follows: first there was ancestor worship, with animism, polytheism, henotheism, and, eventually, theism, following (Marsh, 2013: 363). And where belief in a 'moral high God' preceded or was contemporaneous with ancestor worship or animism, the relevant god, apart from being singular, didn't resemble God in anything like the manner a modern Christian theist would recognize (Marsh, 2013: 358; Boyer, 2001). For many of these high gods weren't particularly "high", at least not morally, and further, they didn't care much for human wellbeing or welfare, whatever one might think of the morality of such divine "neglect" (Marsh, 2013: 358).

Moreover, even if these singular high gods *were* recognizably God – a God modern Christian theists would recognize – ethnographic data suggests that the incidence rate of such beliefs was around fifty percent (Stark, 2007: 61, 62). This is hardly a data point the Christian theist would want to celebrate, especially given the importance of the subject matter involved.

The traditional Christian theist should find the historical time-series distribution of humanities conceptions of God troubling for two reasons. The first is that given God's perfect power and maximal knowledge – he could have guided evolution such that humans formed reliable beliefs about him much *sooner* in their history than they seemingly did. And secondly – and much more importantly – given his perfect goodness, he arguably would have. For there are *very* substantial welfare stakes involved in not knowing that he exists, and that it is *he* that exists.

The first worry – that humans have only recently come to conceive of God more or less reliably (from a Christian perspective) – is more probable on naturalism than theism. For, on naturalism, that (many) humans only recently became monotheists is merely an interesting fact to be explained by anthropology, the cognitive science of religion, or psychology more generally (Marsh, 2013: 361 - 363). On the other hand, many theists – including Plantinga – believe that humans were originally reliable in their knowledge of God (Marsh, 2013: 363; Plantinga, 2000: 177; Genesis 1). In other words, according to them, the first religion was really a reliable monotheism. This is clearly in tension with the historical and contemporary ethnographic evidence. As it is, that humans only recently came to conceive of God more or less reliably should favour naturalism over theism as the better explanation of the time-series data.

Moreover, it doesn't appear that God *had* to use evolution to create humans. For being omnipotent, there were likely other "creative" ways in which he could have done so, ways that didn't result in humans being so historically unreliable with respect to him. On the other hand, there doesn't appear to be many other plausible options available to the naturalist in explaining how humans came to be. If God had other means available to Him in creating humans, and these were such that humans weren't religiously ignorant for most of their earthly tenure, it's strange that he created in the one – or very few – ways that would be naturalistically plausible. Marsh explains:

> [I]f naturalism is true, and if there is to be biological life, we have reason to expect that evolution will be heavily guided by cruel 'survival selection'... and thus reason to expect something like E [i.e. historical *non*-belief in a God that a modern monotheist would recognize]. For when it comes to producing complex structures and establishing them in a population, evolution by natural selection appears to be essential, given naturalism. On the other hand, if theism is true, and if there is to be biological life, the situation looks different. God, being omnipotent, has various creative options. God could perhaps use various non-Darwinian forms of evolution to create. But God could equally use some form of special creation, or some combination of special creation and evolution to create. This creative flexibility would give God a way around E. In light of these claims, then, given the truth of Darwinism, the ratio of the predictive power of naturalism to the predictive power of theism (is higher in this case) (Marsh, 2013: 353, 354).

4.1 Problem 2: Theistic responses to Darwin and the problem of natural non-belief

As noted, the problem of natural non-belief is that it appears – given the evidence that we have – that the 'concept of a High God appears to be a relative latecomer in the cultural history of religion' (Marsh, 2013: 356). It appears that humans didn't have reliable knowledge about God for a long time.

In response, a number of theists have claimed that if one looks at the evidence closely enough, one will see that humanity's first God concept *was* in fact a minimally reliable monotheism. According to Justin Barrett:

> [T]he diversity of (historical) god concepts we see is a consequence of human error and not divine design... (and)... perhaps children would inevitably (in the absence of such original human error) form a perfect concept of God (Barrett, 2009: 97).

And Rodney Stark:

> [I]n primitive times... (God was really there) revealing himself within the very limited capacities of humans to understand (Stark, 2007: 5).

Marsh responds that even if God really *was* present to the ancients, he was evidently more present to some rather than others. For when considering the religious beliefs of different pre-historic or stone-age cultures, Stark's claim that God was adequately present to the ancients, *in general*, appears thin. For, as he notes:

> On Stark's *own* estimation, which is based on ethnographic data for roughly four hundred "pre-industrial" cultures, a huge portion of these cultures (from what I can tell, less than 50 percent) have apparently affirmed a High God, with far fewer affirming an active or moralistic High God that cares about the morality of human beings (Stark, 2007, 60, 61; Marsh, 2013: 358, my emphasis).

Moreover, when one considers examples of what these pre-industrial cultures actually believed, the idea that they had an adequately reliable conception of God becomes much less tenable. For instance, consider the following examples that Pascal Boyer highlights:

> In many places in Africa, (people believe that there are two 'supreme Gods')... but (that) neither of them is really involved in people's everyday affairs, where ancestors, spirits and witches are much more important... (Further)... many spirits are really stupid... In Siberia, for instance, people are careful to use metaphorical language when talking about important matters. This is because nasty spirits often eavesdrop on humans and try to foil their plans. Now spirits, despite their superhuman powers, just cannot understand metaphors. They are powerful but stupid. In many places in Africa it is quite polite when visiting friends or relatives to express one's sympathy with them for having such "ugly" or "unpleasant" children. The idea is that witches, always on the lookout for nice children to "eat", will be fooled by this naive stratagem. It is also common in such places to give children names that suggest disgrace or misfortune, for the same reason (Boyer, 2001: 7, 8).

Moreover:

> [I]n many parts of the world, religion does not really promise that the soul will be saved or liberated and in fact does not have much to say about its destiny. In such places, people just do not assume that moral reckoning determines the fate of the soul. Dead people become ghosts or ancestors. This is general and does not involve a special moral judgement (Boyer, 2001: 8).

At best, the religious beliefs highlighted share a rudimentary resemblance to the God of Christianity – that there is a God, or that the supernatural exists. But apart from these faint congruencies, the evidence strongly suggests that many didn't possess a reliable knowledge of God. On the other hand, the differences between Christianity and these religious beliefs and practices are much more pronounced.

For example, the God of Christianity isn't stupid. He wouldn't be at a loss in understanding metaphor. Indeed, he is infinitely epistemically capable. Secondly, the God of Christianity is not aloof but is believed to be intimately involved in human affairs. Finally, and probably most importantly, one of the most important truths about God is that he loves humanity. And to such an extent that he wishes for everyone to enter into an everlasting and loving relationship with him (John 1; John 3). Hence, I agree with Marsh when he concludes that:

> It can be difficult to imagine that the God of theism is all that present through ideas like these [those that Boyer highlights above]. It can be even more difficult to imagine that cultures that have been restricted to ideas like these experience the kind of valuable divine-human relationship that an unsurpassably loving God would arguably want for them during their lives. Since some present cultures, but especially earlier ones, were so restricted, Stark's suggestion is inadequate (Marsh, 2013: 358).

Another theistic response to the problem of natural non-belief is the claim that religious or supernatural beliefs *are* natural. Justin Barrett:

> Commonly scholars in the cognitive science of religion (CSR) have advanced the naturalness of religion thesis. That is, ordinary cognitive resources [for e.g. agency detection devices, theory of mind capacities, creationist biases, dualist biases, and a tendency to recall and spread minimally counterintuitive or MCI narratives] operating in ordinary human environments typically lead to some kind of belief in supernatural agency and perhaps other religious ideas (Barrett, 2010: 169).

Marsh responds that although the evidence with respect to religious belief may be taken to show that humans are ubiquitously and naturally religious, and thus appears to support the idea 'that God... is doing something to make us religious', it ultimately fails to deflate the problem of natural unbelief (Marsh, 2013: 357). For the problem is *not* that humans aren't naturally religious, but that it appears as if most of them haven't been religious in the right *way*; the available evidence suggests that humanity were *generally* religious, not that they were specifically *Christian*, or even monotheistic. Marsh explains:

> Such a response [that humans are naturally religious and that God is therefore ultimately involved] conflates two very different claims, however. One claim is that God would want people to be religious.

> Another claim is that God would want people to be religious in a particular *way*, namely a way that includes belief and trust in a high moral God in order to enjoy a divine-human relationship. Since theists typically affirm the second, more specific, claim and not just the first, one cannot alleviate the problem (of natural non-belief) by saying that religion in general is cognitively natural or by saying that most people have some sort of religious belief. Put another way, *the problem of natural non-belief is not a problem for religion in general or for supernatural religion in general, but for theism in particular* (Marsh, 2013: 357, my emphasis).

Briefly; given theistic evolution, the problem of natural non-belief should trouble the theist. For the evidence indicates that the development of theistic concepts and beliefs is a recent one. Theists have attempted to defeat this challenge in two related ways.

One, they have argued that humans *are* naturally religious – a claim that has some support from contemporary cognitive science of religion – and thus that God was ultimately involved in making humans religious. Two, some have claimed that God – despite appearances – *was* present to the ancients – albeit not as clearly as to modern monotheists.

Both responses are unconvincing. One, the claim isn't that humans aren't naturally religious, but that they aren't naturally religious *in the right way*. On Christian theism, a central tenet, and a supremely valuable good, is to know that God invites humans to enter into an everlasting 'divine-human' relationship with him. But historical evidence indicates that most of humanity didn't know that such an open invitation was on offer. Two, the claim that God was present to the ancients, despite appearances, becomes much less plausible when one considers the specific religious beliefs different cultures have entertained.

Given all of the above, I think that those theists who believe that God has directed the evolutionary process have as yet failed to provide a plausible response to the problem of natural non-belief. On the other hand, there appears plenty of plausible explanatory real estate within evolutionary naturalism to make sense of the historical distribution of human supernatural belief.[71]

5. Conclusion

In this chapter, the theist's attempts at explaining the evidence with respect to the distribution of human cognitive reliability was discussed. Their general explanatory challenge was cashed out in terms of two specific problems: the epistemic problem of evil and the problem of the hiddenness of God.

The epistemic problem of evil refers to the fact that God guided evolution in such a way that humans often reason in systematically biased ways. This is problematic for several reasons: One, given God's omnipotence, it certainly appears as if he could have created humans that had greater powers of epistemic judgment, and thus bore a closer resemblance to him. And two, given his perfect goodness, one would have expected him to have done so.

[71] See Atran (2002), Boyer (2001), and Wilson (2002) for a range of naturalistically-friendly explanations of human religious diversity.

In reply, it was argued that it's not clear that naturally gifted Bayesians *would* have been the more capable epistemic agents. Moreover, it was shown that it's not a simple matter to determine if society would have been better off were humans epistemically enhanced in this manner. As a result, I don't think the theist should worry too much about the epistemic problem of evil.

However, the problem of the hiddenness of God is another matter. This refers to the fact that the relevant evidence suggests that humans aren't – and for most of their history haven't been – reliable in their knowledge of God. They don't – and didn't – know that he exists and that it is *he* that exists. The acuity of this problem lies in the importance of its subject matter. Indeed, those who don't know God in an adequately reliable manner will not only lose the opportunity to enter into an eternal loving relationship him, but suffer eternally for their ignorance.

Theists have responded to the problem of the hiddenness of God in two important ways. One, they have claimed that those who don't know – the "non-believers" – are always blameworthy for being so unreliable or non-believing. For they argue that God isn't hidden, and has never been. Two, a number of theists have argued that non-belief or cognitive unreliability of this kind may be blameless, but that God has good reasons for allowing it.

The first response – that God is not hidden, and that those who are unbelievers are always blameworthy for their unbelief – was shown to be inadequate. For its untenable to suppose that most of the world's population – past and present – *knowing* that God exists, would be as intellectually dishonest as the historical and contemporary data on the diversity of religious belief indicates. And even if most people *know* that the Christian God exists, appearances to the contrary, it's a stretch to think that so many would willingly decline the offer of entering into an eternally loving relationship with a morally perfect being.

The second response to the problem of the hiddenness of God – that he allows blameless non-belief for good reason – was also shown to be suspect. For there may be plausible reasons that explains why God allows a modicum of religious diversity to obtain – but not in the particular *way* it does. But even if there was such an explanation, I don't see why God would allow humans to be so ignorant with respect to him for so long. In other words, there may be a good reason why the *current* distribution of non-belief appears as it does, but no plausible reason that explains its *historical* distribution.

CHAPTER 5: CONCLUSION

This thesis was divided into two parts. The subject matter of the first – occupying chapters 1 and 2 – was premise 1 of Plantinga evolutionary argument against naturalism (the EAAN). This is the claim that human cognitive faculties would likely be unreliable on naturalism and evolution. In response, naturalists argued that premise 1 is false; there are good reasons to expect human cognition to be trustworthy on the latter. In part two – occupying chapters 3 and 4 – the discussion turned to which of evolutionary naturalism or theistic evolution provides the best explanation for the observed evidence with respect to human cognitive reliability. I claimed that the former is indeed the better (explanation). The importance and relevance of the second question to the first was in showing that if evolutionary naturalism is the better explanation of the two – evolutionary naturalism versus theistic evolution – then evolution, far from being naturalism's kryptonite, is in fact a very close friend.

In Chapter 1, Plantinga's argument in support of premise 1 was discussed. As highlighted, his argument drew its inspiration from the idea that evolution is primarily interested in creature's survival, not the truth-value of their beliefs. Building on this, he argued that there are no naturalistically friendly scenarios on which the link between belief and behaviour would be such that evolution would either be able to select for belief content, or be interested in selecting for mostly true content as it selects for behaviour. As such, he doesn't think that it would be reasonable for the naturalist to expect her cognitive faculties to be reliable.

In Chapter 2, a number of naturalist responses to Plantinga's anti-naturalist argument were considered. I found the most persuasive to be those that traded on the idea that even if evolution has no causal purchase on belief content, the naturalist can still reasonably expect her cognitive faculties to be reliable. For, on functionalism, it was shown that the meaning or content of a belief is given by the role that it plays within a bigger causal context. And hence, as was argued, not just any belief can be attached to any person's behaviour given her desires. Finally, according to the naturalist, she would then be well within her epistemic rights to expect her cognitive faculties to be reliable when this idea is coupled with the plausible assumption that true beliefs would on average be better guides to behaviour than false ones. If this were true, as I think is indeed the case, premise 1 of the evolutionary argument would be false.

In Chapter 3, an overview of the scientific evidence with respect to the reliability of human cognition was presented. A number of proposed evolutionary naturalist explanations of these findings were discussed. The evidence indicated that human sensory-perception is often systematically unreliable in certain environmental contexts. The same was shown to be the case with respect to human reasoning. Finally, the evidence suggests that humans find it hard in coming to reliable concerning subject matters far removed from those salient to their survival. In fact – and of particular importance in terms of this thesis – is human's evident unreliability with regards to "godly" matters.

The evolutionary naturalist explanations I considered all took as their explanatory point of departure the assumption that the human mind should be seen as evolution's answer(s) to natures recurring adaptive challenges. Apart from what I consider to be their general explanatory power, what I found most compelling was their specific

explanation of humans' evident unreliability with respect to the knowledge of God. For, as I made clear, I think it is *here*, if nowhere else, where the contrast in the explanatory power and promise between evolutionary naturalism and theistic evolution is most apparent.

In Chapter 4, I discussed a number of theistic responses to the problems the naturalist thinks the empirical evidence discussed in Chapter 3 raises for the theist. Specifically, I claimed that the theist who believes that God guided the evolutionary process shouldn't worry that humans don't reason in optimally rational ways or that their sensory-perceptual faculties are not as reliable as they probably could have been. For I argued that God may have had good reasons for creating us as we are, reasons that we are not privy to at present. On the other hand, I argued that there is as yet no plausible theistic response to the significant problem that humans evidently don't know much about God, if anything at all.

In short, to the question posed in part 2 of this work, I answer that it may in fact be that theistic evolution is as good an explanation of the evidence discussed as any naturalistic alternative, perhaps even better on some counts. But to my mind, the fact that it misses on the knowledge-of-God data point ultimately makes it the lesser of the two.

The last word. Although Plantinga's argument is ingenious, raising doubts about the expected reliability of human cognition, it ultimately fails. For there are plausible naturalistic reasons to think that human cognition would likely be reliable in a world where naturalism and evolution are true. Further, I think that evolutionary naturalism bests theistic evolution as an explanation of the empirical evidence we have with respect cognitive reliability, especially when it comes to humanity's knowledge of God.

REFERENCES

Ariely, D. 2010. *Predictably irrational: The hidden forces that shape our decisions*.1st edition. New York: Harper Perennial.

Atran, S. 2002. *In gods we trust: The evolutionary landscape of religion.* Oxford, England: Oxford University Press.

Baker-Hytch, M. 2016. Mutual Epistemic Dependence and the Demographic Divine Hiddenness Problem. *Religious Studies* [Electronic], 52(3), August:375-394, doi: https://doi.org/10.1017/S0034412515000359

Barrett, J. L. 2007. Cognitive science of religion: what is it and why is it? *Religion Compass* [Electronic], 1(6), September:768-786, doi: https://doi.org/10.1111/j.1749-8171.2007.00042.x

Barrett, J. 2009. Cognitive Science, Religion, and Theology, in M. Murray & J. Schloss, (eds.). *The Believing Primate: Scientific, Philosophical, and Theological Reflections on the Origin of Religion*. Oxford: Oxford University Press. 76-99.

Barrett, J. 2010. The Relative Unnaturalness of Atheism: On Why Geertz and Markússon Are Both Right and Wrong. *Religion*, [Electronic], 40(3), July:169-172, doi: https://doi.org/10.1016/j.religion.2009.11.002

Benjamin, E. 2019. Principles of Indifference. *Journal of Philosophy* [Electronic], 116(7):390-411, doi: https://doi.org/10.5840/jphil2019116724

Bickle, J. 2020. Multiple Realizability, in Zalta, E. N. (ed.). *The Stanford Encyclopedia of Philosophy* [Online]. Available: https://plato.stanford.edu/archives/sum2020/entries/multiple-realizability/ [2020, June 10].

Blanke, O. & Dieguez, S. 2009. Leaving body and life behind: out-of-body and near-death experience, in S. Laureys & G. Tononi, (eds.). *The Neurology of Consciousness: Cognitive Neuroscience and Neuropathology*. Amsterdam: Elsevier. 303-325.

Boudry, M. 2013. Alvin Plantinga: Where the Conflict Really Lies. Science, Religion and Naturalism. *Science and Education* [Electronic], 22(5), August:1219-1227, doi: https://doi.org/10.1007/s11191-012-9516-y

Boudry, M. & Vlerick, M. 2014. Natural Selection Does Care About Truth. *International Studies in the Philosophy of Science* [Electronic], 28(1), July:65-77, doi: https://doi.org/10.1080/02698595.2014.915651

Boudry, M. Vlerick, M. & McKay, R. 2015. Can evolution get us off the hook? Evaluating the ecological defence of human rationality. *Consciousness and Cognition* [Electronic], 33, May:524-535, doi: https://psycnet.apa.org/doi/10.1016/j.concog.2014.08.025

Bourget, D. & Chalmers, D.J. 2014. What do philosophers believe? *Philosophical Studies* [Electronic], 170, December:465-500, doi: https://doi.org/10.1007/s11098-013-0259-7

Boyer, P. 2001. *Religion Explained: The Evolutionary Origins of Religious Thought*. New York: Basic Books.

Briggs, G.A.D. Butterfield, J.N. & Zeilinger, A. 2013. The Oxford Questions on the foundations of quantum physics. *Proceedings of the Royal Society A* [Electronic], 469(2157), September:1-8, doi: https://doi.org/10.1098/rspa.2013.0299

Cassells, W. Schoenberger, A. & Grayboys, T. B. 1978. Interpretation by physicians of clinical laboratory results. *New England Journal of Medicine* [Electronic], 299(18), November:999-1001, doi: https://doi.org/10.1056/nejm197811022991808

Chakravartty, A. 2017. Scientific Realism, in Zalta, E.N. (ed.). *The Stanford Encyclopedia of Philosophy* [Online]. Available: https://plato.stanford.edu/archives/sum2017/entries/scientific-realism/ [2020, August 11].

Chalmers, D. J. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press.

Childers, G. 2011. What's wrong with the evolutionary argument against naturalism? *International Journal of Philosophy of Religion* [Electronic], 69, October:193-204, doi: https://doi.org/10.1007/s11153-010-9272-3

Churchland, P.S. 1987. Epistemology in the Age of Neuroscience. *The Journal of Philosophy* [Electronic], 84(10), October:544-553, doi: http://doi.org/10.5840/jphil1987841026

Clendinnen, F. J. 1989. Evolutionary Explanation and the Justification of Belief, in *Issues in Evolutionary Epistemology*, K. Hahlweg & C. A. Hooker (eds.). Albany: State University of New York Press. 458-474.

Cosmides, L. & Tooby, R. 1994. Better than rational: evolutionary psychology and the invisible hand. *American Economic Review* [Electronic], 84(2), May:327-332. Available: https://www.jstor.org/stable/2117853?seq=1 [2020, July 16].

Darwin, C.R. 1881. Letter to William Graham. *Darwin Correspondence Project* [Online], Available: https://www.darwinproject.ac.uk/letter/DCP-LETT-13230.xml [2020, September 2].

Dawkins, R. 2006. *The God delusion*. Boston: Houghton Mifflin.

Dretske, F. 1988. *Explaining Behavior: Reasons in a World of Causes*. Cambridge, MA: MIT Press.

Dunitz, J.D. & Joyce, G.F. 2013. Leslie E. Orgel 1927 - 2007: A Biographical Memoir. *National Academy of Sciences* [Online]. Available: http://www.nasonline.org/publications/biographical-memoir=pdfs/orgel-leslie.pdf [2020, March 29].

Ehrsson, H. H. 2007. The experimental induction of out-of-body experiences. *Science* [Electronic], 317(5841), August:1048, doi: https://doi.org/10.1126/science.1142175

Fales, E. 1996. Plantinga's Case Against Naturalistic Epistemology. *Philosophy of Science* [Electronic], 63(3):432-451, doi: https://www.journals.uchicago.edu/doi/abs/10.1086/289920

Fales, E. 2002. Darwin's Doubt, Calvin's Calvary, in *Naturalism Defeated? Essays on Plantinga's Evolutionary Argument Against Naturalism*, J. K. Beilby (ed.). Ithaca, NY: Cornell University Press. 43-58.

Fitelson B., & Sober E. 1998. Plantinga's probability arguments against evolutionary naturalism. *Pacific Philosophical Quarterly* [Electronic], 79(2), December:115-129, doi: https://doi.org/10.1111/1468-0114.00053

Gigerenzer, G. 1991. How to make cognitive illusions disappear: Beyond "heuristics and biases", in W. Stroebe & M. Hewstone (eds.). *European review of social psychology* [Electronic], 2(1), March:83-115, doi: https://doi.org/10.1080/14792779143000033

Gigerenzer, G, & Selten, R. (eds.). 2001. *Bounded Rationality: The Adaptive Toolbox.* Cambridge, MA: MIT Press.

Gilovich, T. Vallone, R. & Tversky, A. 1985. The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology* [Electronic], 17(3), July:295-314, doi: https://doi.org/10.1016/0010-0285(85)90010-6

Gulyás, A. Bíró, J.J. Kőrösi, A. Rétvári, G. & Krioukov, D. 2015. Navigable networks as Nash equilibria of navigation games. *Natural Communications* [Electronic], 6(7651) July:1-10, doi: https://doi.org/10.1038/ncomms8651

Haselton, M. G., & Buss, D. M. 2000. Error Management Theory: A New Perspective on Biases in Cross-sex Mind Reading. *Journal of Personality and Social Psychology* [Electronic], 78(1), February:81-91, doi: https://psycnet.apa.org/doi/10.1037/0022-3514.78.1.81

Haselton, M. G., Nettle, D. & Murray, D. R. 2005. The evolution of cognitive bias, in D. M. Buss (ed.). *The Handbook of Evolutionary Psychology*. New York: Wiley. 968-987.

Haselton, M.G. & Galperin, A. 2011. Error Management and the Evolution of Cognitive Bias [Electronic]. Available: https://psycnet.apa.org/record/2012-10352-004 [2020, April 7].

Howard-Snyder, D. & Moser, P. K. (eds.). 2002. *Divine Hiddenness: New Essays*. New York: Cambridge University Press.

Kahneman, D. & Tversky, A. 1973. On the psychology of prediction. *Psychological Review.* [Electronic], 80(4):237-251, doi: https://doi.org/10.1037/h0034747

Kahneman, D. & Tversky, A. 1974. Judgment under Uncertainty: Heuristics and Biases. *Science* [Electronic], 185(4157), September:1124-1131, doi: https://doi.org/10.1126/science.185.4157.1124

Kahneman, D. Slovic, P. & Tversky, A. (eds.). 1982. *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press.

Kahneman, D. & Tversky, A. 1983. Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review* [Electronic], 90(4):293-315, doi: https://doi.org/10.1037/0033-295X.90.4.293

Kahneman, D. 2011. *Thinking, Fast and Slow*. London: Penguin Books.

Kim, J. 2005. *Physicalism, Or Something Near Enough*. Princeton: Princeton University Press.

Kim, J. 2011. *Philosophy of Mind*. 3rd edition. Boulder, CO: Westview Press.

Ladyman, J. & Ross, D. 2007. *Every Thing Must Go: Metaphysics Naturalized*. Oxford: Oxford University Press.

Law, S. 2011. Plantinga's belief-cum-desire argument refuted. *Religious Studies* [Electronic], 47(2), June:245-256. Available: www.jstor.org/stable/23013384 [2020, February 21].

Law, S. 2012. Naturalism, Evolution and True Belief. *Analysis* [Electronic], 72(1), January:41-48. Available: https://jstor.org/stable/41340792 [2020, March 3].

Lichtenstein, S. Fischhoff, B. & Phillips, L. 1982. Calibration and probabilities: the state of the art to 1980, in D. Kahneman, P. Slovic, & A. Tversky (eds.). *Judgment under uncertainty: Heuristics and biases*. Cambridge, UK: Cambridge University Press. 306-334.

Loewe, L. & Hill, W.G. 2010. The population genetics of mutations: good, bad and indifferent. *Philosophical Transactions of The Royal Society B Biological Sciences* [Electronic], 365(1544):1153-1167, doi: http://doi.org/10.1098/rstb.2009.0317

Loftus, G.R., & Loftus, E.F. 1974a. The influence of one memory retrieval on a subsequent memory retrieval. *Memory & Cognition* [Electronic], 2, May:467-471, doi: https://doi.org/10.3758/BF03196906

Loftus E.F., & Palmer J. 1974b. Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior* [Electronic], 13(5), October:585-589, doi: https://doi.org/10.1016/S0022-5371(74)80011-3

Loftus, E. F. Miller, D. G. & Burns, H. J. 1978. Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology* [Electronic], 4(1):19-31, doi: https://doi.org/10.1037/0278-7393.4.1.19

Loftus, E. F. 1997. Creating False Memories. *Scientific American* [Electronic], 27(3), October:70-75, doi: https://doi.org/10.1038/scientificamerican0997-70

Loftus, E. F. 2005. Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & memory* [Electronic], 12, July:361-366, doi: www.learnmem.org/cgi/doi/10.1101/lm.94705

Maitzen, S. 2006. Divine Hiddenness and the Demographics of Theism. *Religious Studies* [Electronic], 42(2), April:177-191, doi: https://doi.org/10.1017/S0034412506008274

Marsh, J. 2013. Darwin and the Problem of Natural Non-belief. *The Monist* [Electronic], 96(3), July:349-376, Available: https://www.jstor.org/stable/42751257

Mayr, E. 1988. *Towards a new Philosophy of Biology: Observations of an Evolutionist*. Cambridge: Harvard University Press.

McKay, R., & Efferson, C. 2010. The subtleties of error management. *Evolution and Human Behavior* [Electronic], 31(5), September:309-319, doi: https://doi.org/10.1016/j.evolhumbehav.2010.04.005

Melnyk, A. 2003. *A Physicalist Manifesto: Thoroughly Modern Materialism*. Cambridge: Cambridge University Press.

Mercier, H. & Sperber, D. 2011. *The Enigma of Reason.* Cambridge, MA: Harvard University Press.

Mohan, D. M. Kumar, P. Mahmood, F. Wong, K. F. Agrawal, A. Elgendi, M. Shukla, R. Ang, N. Ching, A. Dauwels, J. Chan, A.H.D. 2016. Effect of subliminal lexical priming on subjective perception of

images: A machine learning approach. *PLoS One* [Electronic], 11(2), February:1-22, doi: https://doi.org/10.1371/journal.pone.0148332

Montero, B. & Papineau, D. 2005. A defence of the Via Negativa Argument for Physicalism. *Analysis* [Electronic], 65(3), July:233-237. Available: https://www.jstor.org/stable/3329031 [2020, April 2].

Morin, O. 2016. *How traditions live and die*. Oxford, UK: Oxford University Press.

Myrvold, W. 2018. Philosophical Issues in Quantum Theory, in Zalta, E.N. (ed.). *The Stanford Encyclopedia of Philosophy* [Online]. Available: https://plato.stanford.edu/archives/fall2018/entries/qt-issues/ [2020, August 1].

Ney, A. 2008. Physicalism as an Attitude. *Philosophical Studies*. 138(1):1-15.

Nelson, T. (ed.). 1982. Holy Bible: The New King James Version. Nashville: Thomas Nelson [Online]. Available: https://www.biblegateway.com [2020, July 30].

Nisbett, R. E. & Ross, L. 1980. *Human Inference: Strategies and Shortcomings of Social Judgment*. Englewood Cliffs, N.J.: Prentice-Hall.

Papineau, D. 2020. Naturalism, in Zalta, E.N. (ed.). *The Stanford Encyclopedia of Philosophy* [Online]. Available: https://plato.stanford.edu/archives/sum2020/entries/naturalism/ [2019, November 29].

Petty, R. E. & Wegener, D.T. 1998. Attitude Change: Multiple Roles for Persuasion Variables, in D. T. Gilbert, S. T. Fiske, & L. Gardner (eds.). *The Handbook of Social Psychology*. 4th ed. Boston: McGraw-Hill. 323-390.

Pew Research Center. 2015. *The Future of World Religions: Population Growth Projections, 2010-2050* [Online], Available: https://www.pewforum.org/2015/04/02/religious-projections-2010-2050/

Pinker, S. 1997. *How the mind works.* New York: W.W. Norton and company.

Plantinga, A. 1993. *Warrant and Proper Function*. New York: Oxford University Press.

Plantinga, A. 1994. *Naturalism defeated.* Unpublished manuscript [Online]. Available: https://www.scribd.com/document/143800935/Naturalism-Defeated-Alvin-Plantinga [2020, January 15].

Plantinga, A. 2000. *Warranted Christian Belief*. New York: Oxford University Press.

Plantinga, A. 2002. Reply to Beilby's Cohorts, in Beilby, J. K. (ed.). *Naturalism Defeated? Essays on Plantinga's Evolutionary Argument Against Naturalism*, Ithaca, NY: Cornell University Press. 204-275.

Plantinga, A. 2011a. Content and Natural Selection. *Philosophy and Phenomenological Research* [Electronic], 83(2), January:435-458, doi: https://doi.org/10.1111/j.1933-1592.2010.00444.x

Plantinga, A. 2011b. *Where the Conflict Really Lies: Science, Religion, and Naturalism.* New York: Oxford University Press.

Psillos, S. 2018. Realism and Theory Change in Science, in Zalta, E.N. (ed.). *The Stanford Encyclopedia of Philosophy* [Online]. Available: https://plato.stanford.edu/archives/sum2018/entries/realism-theory-change/ [2020, July 13].

Rabin, M. & Vayanos, D. 2010. The Gambler's and Hot-Hand Fallacies: Theory and Applications, *The Review of Economic Studies* [Electronic], 77(2), April:730-778, doi: https://doi.org/10.1111/j.1467-937X.2009.00582.x

Ramsey, W. 2002. Naturalism Defended. *Naturalism Defeated? Essays on Plantinga's Evolutionary Argument Against Naturalism*, J. K. Beilby (ed.). Ithaca, NY: Cornell University Press. 15-29.

Ridgeon, L. 2003. *Major World Religions*. London: Routledge Curzon.

Ross, C.F., Bohlscheid, J.C. & Weller, K. 2008. Influence of Visual Masking Technique on the Assessment of 2 Red Wines by Trained and Consumer Assessors. *Journal of Food Science* [Electronic], 73(6), August: 279-285, doi: https://doi.org/10.1111/j.1750-3841.2008.00824.x

Russell, B. 1912. *The problems of philosophy.* London: Oxford University Press.

Selten, R. 2001. What is bounded rationality?, in G. Gigerenzer & R. Selten (eds.). *Bounded rationality: The adaptive toolbox*. Cambridge, MA: MIT Press. 1-12.

Sims, D.W. Southall, E.J. Humphries, N.E. Hays, G.C. Bradshaw, C.J.A. Pitchford, J.W. James, A. Ahmed, M.Z. Brierly, A.S. Hindell, M.A. Morritt, D. Musyl, M.K. Righton, D. Shepard, E.L.C. Wearmouth, V.J. Wilson, R.P. Witt, M.J. & Metcalfe, J.D. 2008. Scaling laws of marine predator search behaviour. *Nature* [Electronic], 451, February:1098-1102, doi: https://doi.org/10.1038/nature06518

Spence, C. 2011. Crystal clear or gobbletigook? *The World Fine Wine* [Electronic], 33:96-101. Available: https://www.researchgate.net/publication/284574103_Crystal_clear_or_gobbletigook [2020, July 5].

Stark, Rodney. 2007. *Discovering God*. New York: Harper Collins.

Stevenson, R. J. 2011. Olfactory illusions: Where are they? *Consciousness and Cognition* [Electronic], 20(4), December:1887-1898, doi: https://doi.org/10.1016/j.concog.2011.05.011

Stich, S. 1990. *The fragmentation of reason*. Cambridge, MA: MIT Press.

Stoljar, D. 2017. Physicalism, in Zalta, E.N. (ed.). *The Stanford Encyclopedia of Philosophy* [Online]. Available: https://plato.stanford.edu/archives/win2017/entries/physicalism/ [2020, May 17].

Sudduth, M. 2020. Defeaters in Epistemology, in Fiesen, J. & Dowden B. *The Internet Encyclopaedia of Philosophy* [Online]. Available: https://iep.utm.edu/ep-defea/ [2020, August 12].

Taliaferro, C. & Quinn, L. (eds.). 1997. *A Companion to Philosophy of Religion*. Oxford: Blackwell Publishers.

Theobald, D.L. 2012. *29+ Evidences for Macroevolution: the scientific case for common descent* [Online]. Available: http://www.talkorigins.org/faqs/comdesc [2020, February 14].

Tooley, M. 2019. The Problem of Evil, in Zalta, E.N. (ed.). *The Stanford Encyclopaedia of Philosophy* [Online]. Available: https://plato.stanford.edu/archives/spr2019/entries/evil/ [2020, August 8].

Trouche, E. Sander, E. & Mercier, H. 2014. Arguments, more than confidence, explain the good performance of reasoning groups. *Journal of Experimental Psychology* [Electronic], 143(5):1958-1971, doi: https://doi.org/10.1037/a0037099

Van Doorn, G. Wuillemin, D. & Spence, C. 2014. Does the colour of the mug influence the taste of the coffee? *Flavour* [Electronic], 3(1):10, doi: https://doi.org/10.1186/2044-7248-3-10

Van Fraassen, B. 1996. Science, Materialism, and False Consciousness, in J. Kvanvig. *Warrant in Contemporary Epistemology: Essays in Honour of Alvin Plantinga's Theory of Knowledge*. Lanham, MD: Rowman Littlefield. 149-182.

Vlerick, M. 2012. Darwin's doubt - Implications of the theory of evolution for human knowledge. Unpublished doctoral dissertation. Stellenbosch: University of Stellenbosch. Available: https://eur03.safelinks.protection.outlook.com/?url=http%3A%2F%2Fhdl.handle.net%2F10019.1%2F71595&data=02%7C01%7C%7C5c1824b0f09c44e75b0108d85af2ff25%7C84df9e7fe9f640afb435aaaaaaaaaaaa%7C1%7C0%7C637359347002176266&sdata=tD7Egm9rZETQRKThQD2I6y2480aKkGssuygZL629Hlk%3D&reserved=0

Wason, P. C. 1966. Reasoning, in B. M. Foss (ed.). *New horizons in psychology*. Baltimore: Penguin Books. 135-151.

Wason, P. C. 1968. Reasoning about a rule. *Quarterly Journal of Experimental Psychology* [Electronic], 20(3), April:273-281, doi: https://doi.org/10.1080%2F14640746808400161

Wilson, D.S. 2002. *Darwin's cathedral*. Chicago: University of Chicago Press.

Wilson, J. 2006. On Characterizing the Physical. *Philosophical Studies* [Electronic], 131, October:61-99, doi: https://doi.org/10.1007/s11098-006-5984-8

Witmer, G.D. 2012. Naturalism and Physicalism, in R. Barnard & N. Manson (eds.). *Continuum Companion to Metaphysics*. London: Continuum International Publishing Group. 90-120.

Wright, R. 2009. *The Evolution of God*. New York: Little, Brown and Co.